Is the Population Census Useful to Understand the Pedestrian Friendliness of the Built Environment?

An Empirical Examination in Portland

by

Hannah M. Young

A Masters Project submitted to the faculty

of the University of North Carolina at Chapel Hill

in partial fulfillment of the requirements

for the degree of Master of Regional Planning

in the Department of City and Regional Planning.

Chapel Hill

May 2005

Approved by:

______________________________ ______________________________

READER (optional) ADVISOR

Acknowledgements

I am very grateful to Professor Daniel Rodríguez for his initial help selecting a relevant topic and for his continued insightful guidance and encouraging feedback. Professor Yan Song has provided kind assistance with GIS and as well as the Portland data from the RLIS Lite CD (2003). I have much appreciation for Lisa Crooks' invaluable help solving numerous GIS dilemmas.

Special thanks to Robert Schneider (Toole Design) for initial feedback on the BEI, Michael Greenwald, Ph.D. (University of Wisconsin) for the PEF shapefiles, Marc Schlossberg, Ph.D. (University of Oregon) for sidewalk files, and Mark Bosworth (Metro) for historic Portland data. Support for this project was provided in part by a grant from The Robert Wood Johnson Foundation®, Princeton, New Jersey.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1. ABSTRACT

This study proposes a built environment index (BEI) that is useful to researchers assessing the quality of the built environment on factors related to walking, bicycling, and transit use. The BEI is an accessible tool because it utilizes readily available U.S. Population Census data and employs a simplified GIS analysis. The index has numerous applications in studies related to health, mode choice, and land use policy. Similar indices have been used in travel models to enhance model accuracy and for the targeting of priority funding areas for cost-effective infrastructure investment. In this study we review previous research utilizing similar indices to examine the types of variables that ought to be included in the index.

We compare three different methods for calculating the BEI and conclude that two methods are particularly robust. We then offer a comparison of the BEI with the well-studied pedestrian environment factor (PEF) from Portland. This comparison, however, is tenuous and suggests that though there may be similar uses of the indices, they capture different concepts. We conclude that the BEI is a powerful tool to gain a broad overview of the built environment. It is flexible to manipulation and can be tailored for specific research questions as well as supplemented with more fine-grained approaches where a neighborhood scale analysis is required.

2. INTRODUCTION

Interest in the relationship between transportation and land use has led to the examination of the attributes of the built environment associated with walking and bicycling. Because numerous attributes of the built environment are expected to be associated with behavior, researchers have developed a composite score or scores summarizing relevant dimensions of the environment. These indices provide a tool for combining many characteristics of the urban form into a single variable or index.

There are numerous research and planning uses for indices or composite scores of the built environment. First, transportation and land use planners have a growing interest in pedestrian, bicycle, and transit modes. As a result, they have begun incorporating variables associated with such alternative modes into transportation models. This incorporation creates the first step towards recognizing these modes as legitimate means of transportation.

Interest in air quality and pollution reduction (Clean Air Act Amendments of 1990 and Intermodal Surface Transportation Efficiency Act of 1991) has focused policies and initiatives towards supporting non-auto modes. More recently, concern for public health has placed attention on the connection between physical activity levels and the built environment (Handy et al., 2002; Sallis et al., 2004). Inclusion of built environment variables in travel models ultimately allows for the testing of policies that affect the land use travel connection and potentially address these social objectives. For this reason, researchers have also used built environment indices to develop sampling strategies across various environments. SMARTTRAQ in Atlanta is one of the first large-scale data collection programs relying on indices of the built environment (Sallis et al., 2004).

A second use for built environment indices is demonstrated by the pedestrian and bicycle friendliness index (PFI) created by the Maryland-National Capital Park Planning Commission (M-NCPPC) in the late 1980s. This index was used in conjunction with the County's adequate public facilities ordinance (Replogle, 1990), intended to restrict growth to regions where infrastructure could support development. The index became important for measuring the current capabilities of an area with regards to transportation and for the establishment of priority funding areas.

Finally, built environment indices, as discussed by Ulmer and Hoel (2003), can be used to guide real estate location decisions. An index provides homebuyers with an array of built environment attributes to guide them in making the optimal location choice. Additionally, the index could be used in the selection of areas qualifying for Location Efficient Mortgages.

Despite the increasing use and importance of built environment indices, several questions remain. First, several methods of calculating such indices have been used. This has limited the comparability of indices across studies. Second, many of the indices have been developed with a substantial commitment of resources. Expert planners or consultants were asked to perform site visits and use their judgment in rating various environments. This approach not only raises questions of reliability, but also involves heavy resource investments. Few communities have the luxury of devoting valuable resources to such a task. Third, the increasing availability of geo-referenced data allows for alternative indices of the environment that may pose as substitutes to their labor-intensive counterparts. In this context, the current study makes three contributions: a) It develops a twelve-factor built environment index (BEI) relying mostly on U.S. Population Census data; b) It determines the reliability of three common methods of developing indices of the built environment in the particular context of Portland, Oregon; and c) It compares the reliability of the BEI with the pedestrian environment factor (PEF) created by Portland Metro in the early 1990s.[1] The PEF was one of the earlier indices created and it has been broadly studied and utilized, even in several other locations (Eash, 1997; Greenwald and Boarnet, 2001; Hess and Ong, 2001; DKS, 2002). Within Portland's travel demand models, the PEF has been shown to have predictive validity, improving model accuracy by providing a means to capture the influence of environmental factors on travel choices (Parsons, Brinckerhoff, Quade, and Douglas, Inc. et al. 1993; Cambridge Systematics, Inc. et al., 1992).[2]

This paper begins with a discussion of built environment indices. Particular detail is provided regarding the PEF index and its use in various studies. Second, other indices are examined and their contribution to the understanding of what variables are important and the types of applications in which indices are useful is explored. Third, an extended literature review provides further context for the discussion of variables associated with alternative transportation. The fourth section, *Methods*, describes the data sources used for the creation of the BEI, a brief

---

[1] Metro is the regional government for the Portland metropolitan area.

[2] Due to dissimilarities between the types of factors included in the index, the method, and the approach, it is not anticipated that the PEF and BEI will have identical outcomes for each location. However, it is expected that there will be a similar ranking order between the two indices (places with higher PEFs will have higher BEIs). Additionally, a subset of the BEI, the pedestrian and bicycle infrastructure domain, may compare more closely with the PEF as it captures a restricted set of related concepts.

summary of the GIS steps taken to process the data, and a description of the three different methods employed to calculate the index (principal components analysis, cluster, and naïve ranking). Fifth, the results section provides a discussion of the descriptive statistics for these three methods, compares the three approaches with each other, and compares the approaches with the PEF. Finally, conclusions are drawn regarding the implications of the comparisons and the applications of the BEI.

3. BUILT ENVIRONMENT INDICES

3.1. THE PEDESTRIAN ENVIRONMENT FACTOR (PEF)

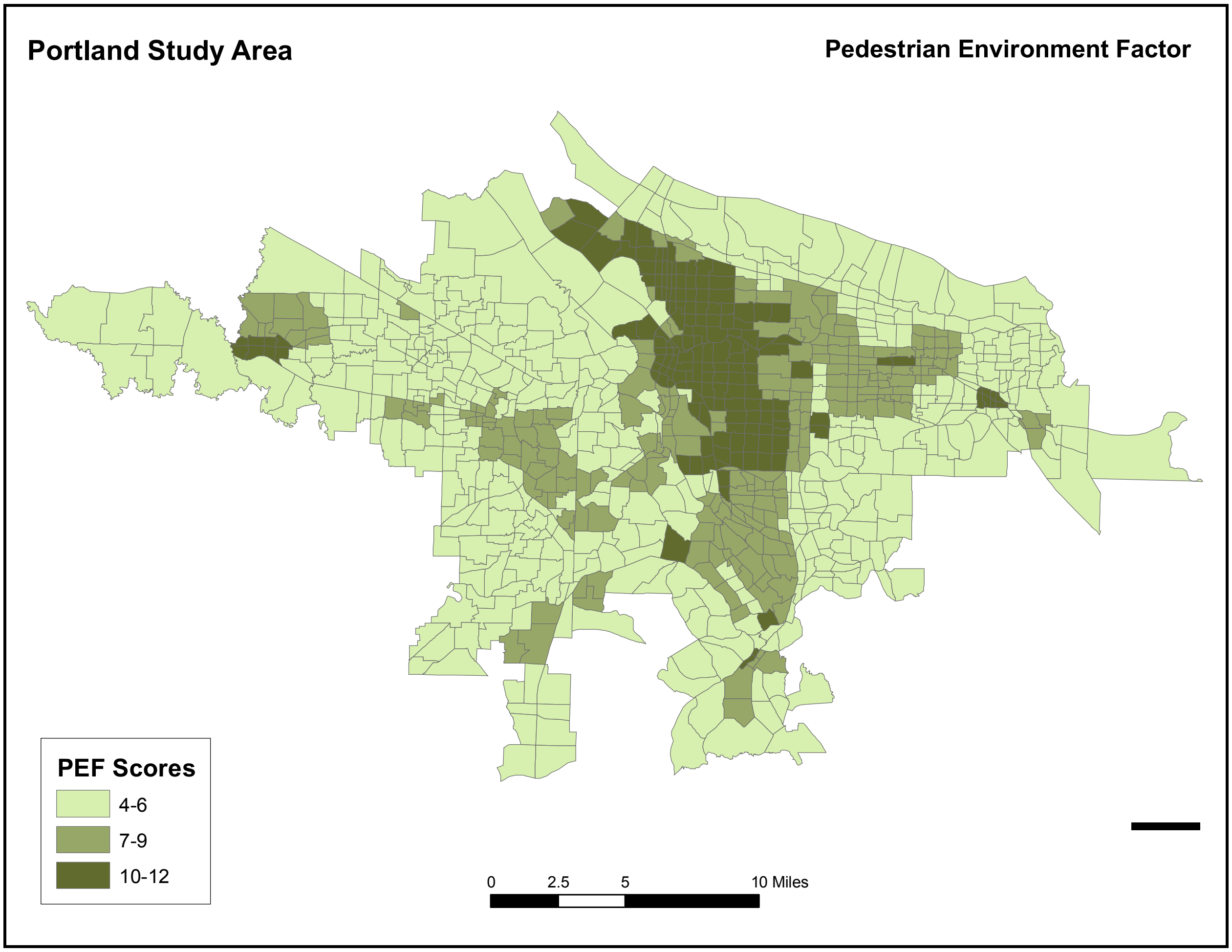
The first built environment index discussed in this paper is Portland's pedestrian environment factor (PEF), which is compared with the BEI in the *Results* and *Discussion* sections. The Land Use, Transportation, Air Quality (LUTRAQ) Connection study published in the early 1990s utilized this composite index in the analysis of a projected new highway and alternatives in the Portland region. This was correlated with the 1985 home interview survey of travel behavior and was used to improve the accuracy of regional travel forecasting models. Higher PEF scores (better pedestrian environments) were clearly correlated with increased non-auto mode choice (Parsons, Brinckerhoff, Quade, and Douglas, Inc. et al., 1993). Additionally, two linear regression models that predict household vehicle trips and vehicle miles showed PEF is a statistically significant factor in the reduction of vehicle trips and miles.

The PEF consists of the following four factors:

1. Ease of street crossings (evaluation of key intersections based on width, extent of signalization, and traffic volumes)
2. Sidewalk continuity (evaluation of extensiveness of sidewalks on principal arterials served by transit, or likely to be served by transit, and extent of sidewalks on neighborhood collector streets)
3. Local street characteristics (evaluation of connectivity of street system, grid vs. cul de sac)
4. Topography (evaluation of extensiveness of sloping terrain and steepness of slopes)

A team of four staff members assessed and scored each of the four factors for key intersections across the Portland region. Each factor was assigned a value from one to three and the results were summed for a total score between four and twelve (a higher score indicates a better walking environment; see Figure 1). The team used a simplified Delphi approach to achieve consensus on the scores. This process was repeated throughout the Portland metro region and a score was assigned for each of the 400 traffic analysis zones (TAZs).

Figure 1. PEF Scores by TAZ for Project Study Area.



The study then matched the zonal PEF scores with travel data from over 15,000 trips for more than 2,400 households. They found that households in TAZs with higher PEF scores make fewer trips by auto and a greater percentage of trips by transit, bicycle, and walk modes. Following this correlation of PEF scores and travel patterns, two linear regression models, as mentioned above, were created with the goal of isolating the individual affects of land use variables in order to test their assumed effect upon vehicle trips and vehicle miles (Parsons, Brinckerhoff, Quade, and Douglas, Inc. et al., 1993). The first model, vehicle miles traveled (VMT), explains 25% of the observed variation in household VMT. A single unit increase in the PEF score lowered the household VMT by 2.5%. On the other hand, a two-point increase in PEF score was estimated to cause a reduction of 455 miles per year in annual household VMT or 182 vehicle miles per person, a rather large effect when calculated at a regional scale. Similarly, the second model, household vehicle trips, explained 23% of the data and showed that increases in PEF scores were associated with decreases in vehicle trips. Specifically, an increase in PEF score from four to seven decreased vehicle trips by 7.3%, while an increase in PEF score from seven to ten decreased vehicle trips by 3.6%.

As part of the LUTRAQ study, three utility models were modified by the addition of the PEF and other land use density and heterogeneity variables in order in improve their accuracy (Cambridge

Systematics, Inc. et al., 1992). The first model, auto ownership, estimates household car ownership levels. As might be expected, the utility of owning a vehicle decreases in areas with increasing PEF. The use of PEF in the model improved the model's predictive capabilities at the extreme high and low ends of PEF scores.[3] The second model, pre-mode choice, predicts the percentages of travel using walk/bicycle modes for origin-destination pairs. The PEF was a significant variable in this model and an increase in either PEF or retail density, another model variable, increased the probability of walk/ bicycle trips. The third model, mode choice, estimates the number of auto versus transit trips. Here, the PEF also positively influenced transit mode choice.

In a more current study, Greenwald and Boarnet (2001) use the PEF within an index that examines the impact of design upon non-work walking travel trips. This study uses data from the 1994 Portland Travel Diaries (a two-day travel record). Their model (ordered probit regression with non-work trips as the dependent variable) uses a variety of variables including neighborhood variables, regional land uses, and socio-demographic characteristics of participants. Neighborhood variables include the PEF, percent area within a quarter-mile buffer covered by grid streets, population density, and retail employment density. Regional variables include population density and retail density for the zip code. Density had a localized effect on walking but not much impact at the regional level. They conclude that although it is difficult to influence non-work mode choice, it can be impacted indirectly through the time cost of travel at the neighborhood level. Supportive urban forms for pedestrian trips include neo-traditional design, retail and employment centers within a fifteen-minute walking distance from housing, high population density, and rectilinear streets.

Recently, Metro updated their travel model to a composite urban design measure (Lane Council of Governments, LCOG, 2003). After analyzing eight different factors, they concluded that a single measure that captured several aspects of accessible mixed-used development best fit the data.[4] This measure combines three types of density factors including density of local street intersections, density of households, and density of retail business. The combined variable was significant and explained the data better than individual elements of the measure. While the variable impacts mode choice, it does not affect auto ownership.

3.2. OTHER BUILT ENVIRONMENT INDICES

Many other indices have been created using different combinations of various factors. This section reviews five different studies to examine the variables utilized, the applications of the index, and the outcomes of the studies. Each study takes a different approach, but common themes emerge regarding the importance of mixed land uses (especially certain types of retail and commercial in proximity to residential), high population density, grid street patterns, and other neighborhood design elements.

[3] The original model was compared to the enhanced model and survey results.

[4] The eight factors considered in the study include skinny (narrow) streets, census blocks, dissimilarity index, mean entropy, building coverage index, less auto-dependent urban form (composite variable for density, land use mix, and circulation), residential parking permit areas (on-street parking restricted by permit), and business establishments.

In a study that utilizes a variation of the PEF, Eash (1997) queries how to reduce auto dependence through urban design. He proposes that the key factor is the placement of activities and transit within walking distances. Design can increase transit ridership through reducing origin and destination distances from transit stops. Land use is a tool planners can use to bring these places together and thereby shorten trip distances, which will allow use of alternative modes. Additionally, placing retail and commercial closer to residential can modify trip generation. The Chicago Area Transportation Study (CATS) used a PEF based on the number of census blocks in a quarter section. This variable, along with an auto-work trip mode share variable is used in the logit ownership model to predict the utility of household vehicle ownership. Higher PEF scores were associated with older or more urban neighborhoods.

The Maryland-National Capital Park Planning Commission developed a pedestrian and bicycle friendliness index (PFI), which is similar to the PEF in the method of data collection (field study). The validity of this index was tested in a home-to-work nested logit mode choice model. The index is composed of the following five factors: the amount of sidewalks, land use mix, building setbacks, transit stop conditions, and bicycle infrastructure. The scores for each were added for a total score ranging from 0 to 1, where 1 is the most pedestrian friendly environment. Theses scores were assessed by consultants and are therefore subjective measures. In the mode choice model, the index indicated that a 0.1 increase in index score resulted in a one-minute increase in auto out-of-vehicle time for the main mode choice model (auto vs. transit). In the walk access submodel of transit mode choice, the index indicated that a 0.1 increase in the index added five minutes to acceptable travel times for walking (Cambridge Systematics, Inc. and Barton Aschman Associates, 1994).

Krizek (2003) synthesizes over 50 studies and creates a new index, focusing on neighborhood accessibility variables. Neighborhood accessibility is closely tied to pedestrian friendliness and is distilled into the three major concepts of density, land use mix, and streets/ design. Density should be a joint measure of jobs and housing since it is the mix of the two factors that is critical. Land use mix can be measured in one of the following three ways: inspection or survey of the area, analysis of employment data, and calculation of entropy and dissimilarity indices. As summarized by Krizek, not only are certain land uses valued more highly in proximity to residential uses (Banerjee and Baer, 1984) but also the spatial and functional complementation of land uses is important (Lynch, 1981). Krizek categorizes the third component, streets/ design, into four sections: street patterns, pedestrian amenities, experimental elements, and composite indices. Street patterns examined include grid or four-way versus three-way intersections. Experimental elements aim to capture the quality of the pedestrian area; concepts that are hard to quantify but important to pedestrian comfort include speed of traffic, width of sidewalk, and tree cover.

Krizek applied his index of neighborhood accessibility to the Central Puget Sound region in Washington. He measured population density, housing unit density, number of employees (for retail, food, restaurants), and block size (intersections/ area) using 150-meter grid cells as the units of analysis. The key to this index is that the value of each grid cell is influenced by the surrounding cells (the grid score is the average of grid cells within a quarter-mile radius). Factor analysis was then used to assign final values to each grid. The validity of the model was tested comparing the variables listed above with a subjectively assigned accessibility score derived from a quasi-field study for 70 sample neighborhoods. The index variables were used as the independent variables in a regression analysis and the scores assigned by neighborhood reviewers were used as the dependent variable. Overall, the accessibility variables explained about 73% of the variation in subjective scoring, indicating a high degree of similarity between the index and field-study approach.

Ulmer and Hoel (2003) created a Pedestrian and Cycling Accessibility Measure (PCAM), which relies on the "three D's" concept of density, diversity, design (Cervero and Kockelman, 1997). Their index is very useful because it can be used to target infrastructure improvement programs. This usefulness is based the three-component nature of the index, which allows each component to be analyzed individually. For example, the first two components, density and diversity scores, suggest locations where infrastructure improvements would have the greatest impacts (areas of higher density and greater diversity). Where the design component indicates that the infrastructure in these locations is particularly lacking, the area should be targeted for investments.

The first component of the index, density, contains measures of the proximity of destinations to residential locations as measured by the weighted density of desirable destinations within a mile of residential locations for walk trips or three miles for bike trips.[5] Destination is particularly important to walk/ bike trips because these trips are relatively short and require nearby locations. The second component, diversity, assesses the mix of destinations within a quarter-mile radius of another destination. Both density and diversity components receive points for population density and employment densities above a certain threshold. The third component, design, is a measure of pedestrian and bicycle infrastructure based on the road network, sidewalks and bike lanes, business of street, and street crossing at busy roads. Ulmer and Hoel (2003) apply this index to the Charlottesville/ Albemarle region in Virginia and explore meaningful ways to display the results.

Frank et al. (2004, under review) utilized an index of density, mix of land use, and connectivity to understand factors causing physical activity. Study participants in King County, Washington and the Baltimore-Washington, DC region wore an accelerometer that recorded their number of minutes of moderate-to-vigorous physical activity per day. The results were correlated with the multi-dimensional walkability index. This index measured net residential density (residential units per area of residential land use per block group), retail floor area ratio (retail building area per retail land area), land use mix (entropy score), and intersection density (intersections with three or more segments per acreage of block group). The index formula was derived from the sum of the Z scores where intersection density was twice the weight of the other two factors. This formula was derived from correlating different weighting schemes with survey results so that the highest index scores appeared in areas where the highest walking was recorded. Site visits correlated the scores with actual perceptions of the area as a confirmation of the index. Thirty-two neighborhoods were matched on walkability, income levels, and demographics. In both geographic regions they observed greater walking and transit use in high-walkability neighborhoods and more single-occupancy vehicle use in low-walkability neighborhoods.

[5] Desirable destinations include drugstore, grocery store, library, post office, small food, park, bank, dry cleaner, beauty/ barber shop, school, friend's place, restaurant, and place of work (Banerjee and Baer, 1984).

4. REVIEW OF RELEVANT STUDIES

There is currently no commonly accepted list of the variables that ought to be included in an index and the types of variables examined depend in part on the application of the index. This section reviews the empirical evidence on the relationship between transportation and the built environment in order to identify commonalities across studies and inform the choice of variables for the BEI.

Travel is perceived as a derived demand. It is not necessarily something that people desire, but they engage in it for the purpose of reaching desired activities. Cervero and Kockelman (1997) use factor analysis to test New Urbanist assumptions that mixed land use, density, and diversity in urban design lead to more pedestrian trips. They examine several variables related to the "three D's". Density is measured for population and employment, and is represented by accessibility of jobs, which serve as a proxy for proximity and compact land use. Diversity includes a dissimilarity index[6], entropy[7], vertical mixture (parcels having commercial, retail, and other land uses), intensities of land use, activity center mixture (entropy within commercial areas), commercial intensities (ratio of developed areas to commercial uses), and proximity of developed areas to commercial-retail uses. The last "D", design, measures streets (patterns, coverage, speeds), pedestrian and cycling amenities (sidewalks, traffic signals, bike lanes), and site design (on-street and off-street parking for commercial/ retail parcels). After controlling for socio-demographic variables they find that gird-iron street patterns, as measured by four-way intersections and low on-street parking near commercial sites, show less non-work related single-occupancy vehicle trips. They also conclude that neighborhood characteristics are more predictive of non-work trip mode choice than mode choice for work-related trips.

In a study focusing on factors impacting auto-ownership, Hess and Ong (2001) examine the pedestrian environment, land use mix, and proximity to transit and light rail. They take an alternative approach and focus on auto ownership as a primary factor in mode choice. They model the vehicle ownership decision (number of cars) in an ordered logit regression. The actual factors included in the study are household characteristics (household income and size, dummy variables for single family home, white householder, male householder, non-senior householder), neighborhood characteristics (median income, household density, percentage white), and design characteristics (dummy variables for land use mix and pedestrian environment factor where 1 equals diverse mix/ good pedestrian environment, 0 equals otherwise). They also included transit access (dummy variable of 1 for good accessibility, 0 otherwise) and light rail access (dummy variable of 1 for households located within a quarter-mile of light rail, 0 otherwise). This study looks at the Portland area and uses data from the Oregon and Southwestern Washington 1994 Activity and Travel Behavior Study. Their model indicates that land use mix has a significant impact on owning a car, but individual variables are not significant independently.

Ewing and Cervero (2001) review over 50 studies linking travel demand and the built environment. They cluster studies into five different classes based upon the type of research:

[6] The dissimilarity index is the proportion of dissimilar land uses in a grid cell. The index increases as the number of dissimilar adjoining uses increases.

[7] Entropy involves the comparison of an area to the surrounding region. A place is maximally entropic if the proportion of land uses is the same between sub-region and region.

comparison of neighborhoods and activity center designs, testing of land use variables, testing transportation network variables, testing urban design variables, and studies using a composite index.

While the focus of this project is on composite indices, Ewing and Cervero's review of studies testing land use variables is particularly informative to the discussion of key variables in mode choice. General themes emerge from their study summaries. Density (usually measured by employment and households) is a key variable in a large number of studies and correlates with increased transit and pedestrian trips and decreased VMT. However, density itself may not be the determining factor but may serve as a proxy for other variables that encourage alternative modes, such as grid pattern, proximity of mixed uses, or dis-utilities such as high parking costs and congestion. The type of land use also plays a part in determining mode shares. Mixed use has been studied broadly and correlated with transit share increases. Retail and commercial, in particular, are important uses related to transit and walk/bike trips. Regional access, the concept of inter-neighborhood connectivity, is also associated with reduced miles traveled.

In a study focused on the physical activity connection of land use and transportation, Sallis et al. (2004) analyze Atlanta travel data from 2000 and determine that many trips are a short enough distance to allow for the selection of non-auto modes, yet these modes are not utilized to the extent possible. This begs the question, what other factors play a role in non-auto mode choice when distance is not the prohibitive factor? Another factor related to the impact of distance on physical activity is the location of parks and recreational facilities proximate to homes. The presence of such facilities near a residence is positively correlated to adult physical activity. Kitamura et al. (1997) also looked at park density and found transit and walk/bike trip shares increased with closer proximity to parks (as reviewed in Ewing and Cervero, 2001). This is of particular interest to this project as park density is one of the components of the BEI. Even for children, this factor appears to be significantly related to activity levels (Sallis et al., 1993 and Blommeart et al, 1981 as mentioned in Sallis et al., 2004). As agencies and professionals focus on the synergy between the urban form and physical activity levels, built environment indices will be a valuable research tool for measuring variables potentially impacting active transport.

Schlossberg and Brown (2004) analyze several different factors in a GIS analysis of transit oriented development (TOD) neighborhoods in Portland. They provide an analytical GIS framework for the classification of neighborhoods and thereby illustrate disparities between the qualities of pedestrian environments across TODs. Though not entirely unexpected, their results show that some places designated as a TOD may not meet commonly accepted TOD criteria when the built environment is rigorously analyzed. The dataset captures the concepts of network classification (ordering of streets based on how pedestrians use them), pedestrian catchment areas (analysis of the actual network versus theoretical five-minute or ten-minute walk), and impedance-based intersection intensities (the concentration of intersections with pedestrian choice such as three or four-way intersections versus dead-ends and no-choice options such as larger arterials). They examine twelve factors (six separate factors, with high and low measures of each) which include quantity of accessible paths, quantity of impedance paths, pedestrian catchment area ranking, impedance pedestrian catchment area ranking (pedestrian catchment areas with major impeding roads removed from the network), intersection density, and density of dead ends. This paper shows that such a fine-grained analysis (performed here at quarter-mile and half-mile distances) provides insights into the pedestrian response to the built environment that cannot be assumed based on the place type designation alone (e.g. TOD).

Though many of these studies have shown correlations between land use variables, density, design variables and travel behavior, it is difficult for studies to address the issue of causality, which leaves the question of the impact of policy unanswered. One explanation is that people who live in more dense or mixed use neighborhoods self-select through their location decisions and prefer to walk or take transit. In this case the built environment is not inducing them to use alternative modes, although it certainly may be facilitating their mode choice. Thus far research has demonstrated that there is at least latent demand for housing having greater accessibility.

Handy et al. (2005) attempted to address this issue of self-selection and causality. In a cross-sectional study examining eight Northern California neighborhoods, they demonstrate that attitudes play a large role in determining if a person takes alternative modes of transportation. In fact, the built environment had a minimal impact on travel behavior of individuals in this cross-sectional study. Instead, pro-bike/walk and pro-transit attitudes were significantly correlated with reduced driving. However, in another part of the study, they performed a quasi-longitudinal assessment of travel and attitudes for recently moved households versus longer-time residents. Accounting for attitudes, the results suggest that travel behavior changes are strongly associated with changes in the built environment. Of the variables analyzed, increases in accessibility had the greatest impact on reduced driving. While this study does not completely answer the question of causality, it suggests that policies can indeed impact travel behavior.

5. METHODS: DEVELOPMENT OF THE BUILT ENVIRONMENT INDEX (BEI)

The BEI is composed of twelve different factors, which are divided into three domains (Table 1). These factors are aggregated values from the Census block group or tract and will be further aggregated (or disaggregated in some cases) to the TAZ level. They represent the range of land use variables associated with non-auto transportation. The first domain, development intensity factors, is composed of population density, housing unit density, employment density, and park density. While these factors clearly capture the concept of density, they also broadly represent other important concepts such as jobs/ housing balance and land use mix. The second domain, motorized transportation factors, contains roadway density, bus route density, transit commuting, proximity to subway station, and proximity to commuter rail (the last factor is not applicable to this study). These factors contain elements of accessibility, connectivity, and mobility concepts. These principals are also found in the third domain, pedestrian and bicycle infrastructure factors, which includes sidewalk density, sidewalk coverage, and pedestrian and bicycle commuting factors. The commuting percentages (transit and pedestrian/ bicycle) provide a creative link between other environmental factors and behavior.

5.1. DATA SOURCES

The spatial scope of analysis initially included the 1,247 traffic analysis zones (TAZs) in the Portland Metro area (Figure 2). This boundary is coincident with the boundaries of the three Oregon counties of Clackamas, Multnomah, and Washington and Clark County, Washington. Data constraints limited the study area boundaries because information for Clark County, Washington was not available on the RLIS Lite CD (see comments below). The spatial extent of the project study area was further limited by the sidewalk coverage data.[8] This data did not cover TAZs at the periphery of the metro area. The final study area included 873 TAZs (Figure 3).

Table 1. Components the BEI.

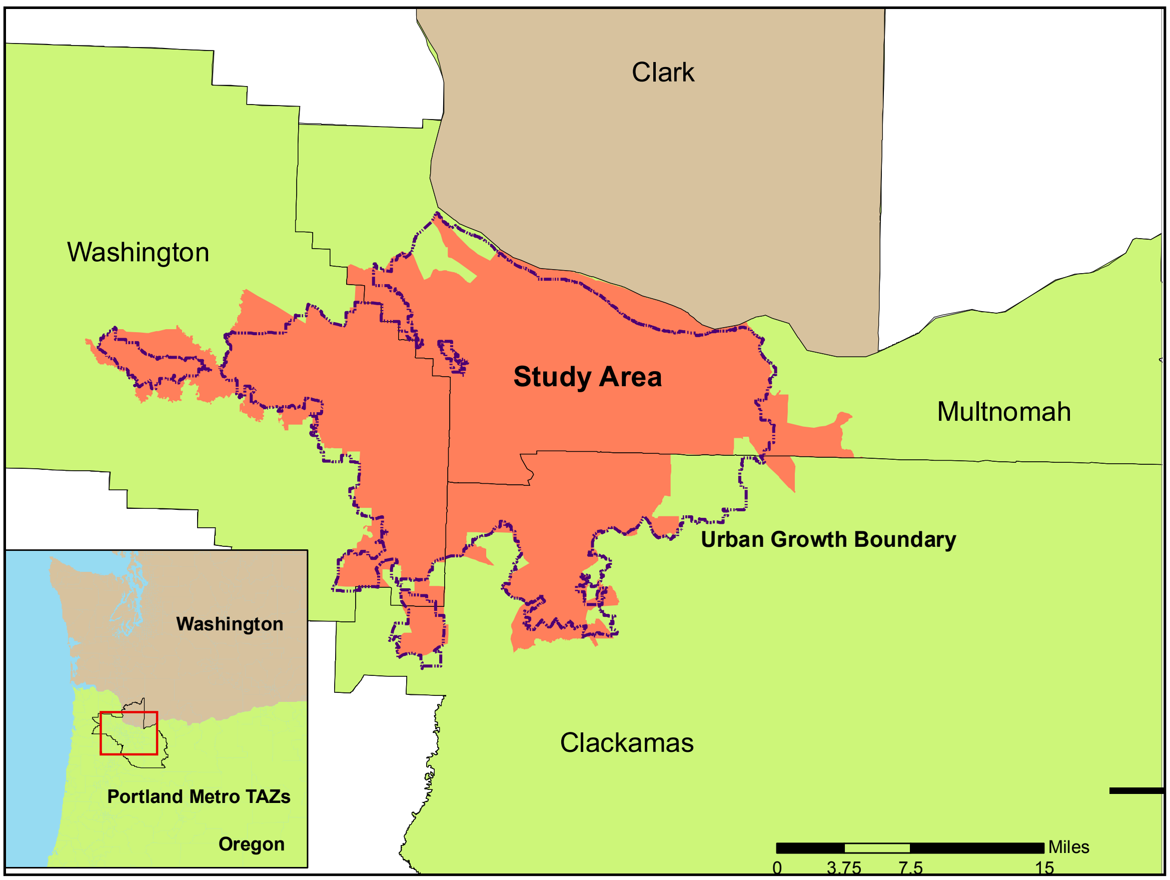
Factor	Measure	Source	Year	Scale
Domain 1: Development Intensity Factors				
Population density	Population/gross acre	Census	1990	Block Group
Housing unit density	Housing units/gross acre	CTPP	1990	Tract
Employment density	Jobs/gross acre	CTPP	1990	Tract
Park density	Parks/gross acre	Portland Metro	1990	NA
Domain 2: Motorized Infrastructure Factors				
Roadway density	Roadway centerline miles/gross acre	TIGER files	1990	NA
Bus route density	Bus route centerline miles/gross acre (includes overlapping routes)	Portland Metro	1996	NA
Transit commuting	Percent commuting by public transit mode	Census	1990	Block Group
Proximity to subway station	Metro rail station within 0.5 miles or 1 mile of the CAZ center	Portland Metro	1990	NA
Proximity to commuter rail	MARC rail station within 0.5 miles or 1 mile of the CAZ center	Not a relevant factor in Portland.		
Domain 3: Pedestrian and Bicycle Infrastructure Factors				
Sidewalk density	Sidewalk miles/gross acre	Portland Metro	2002	NA
Sidewalk coverage	Sidewalk miles/roadway centerline miles	Portland Metro	2002	NA
Pedestrian and bicycle commuting	Percent commuting by pedestrian or bicycle mode	Census	1990	Block Group

[8] Sidewalk coverages are available from Portland Metro but were not provided on the RLIS Lite CD.

The four sources of data used in this project include 1990 Census data found on Census CD + Maps, Release 4.0, from GeoLytics, the 1990 TIGER files, the 1990 Census Transportation Planning Package (CTPP) found online, and the data from Metro (RLIS Lite CD and historic data). See Table 1 for the list of sources for each of the eleven factors.

The Census data used in this project originated from three different files, as mentioned above. Census 1990 data included population and commute percentages for block groups. Road data was obtained from the 1990 TIGER files. CTPP 1990 data obtained for the project included housing units from Part I, the residential section of the package and employment from Part II, the work section. The tract level was the smallest unit of analysis available for Portland area CTPP data.

Figure 2. Portland Study Area Overview.

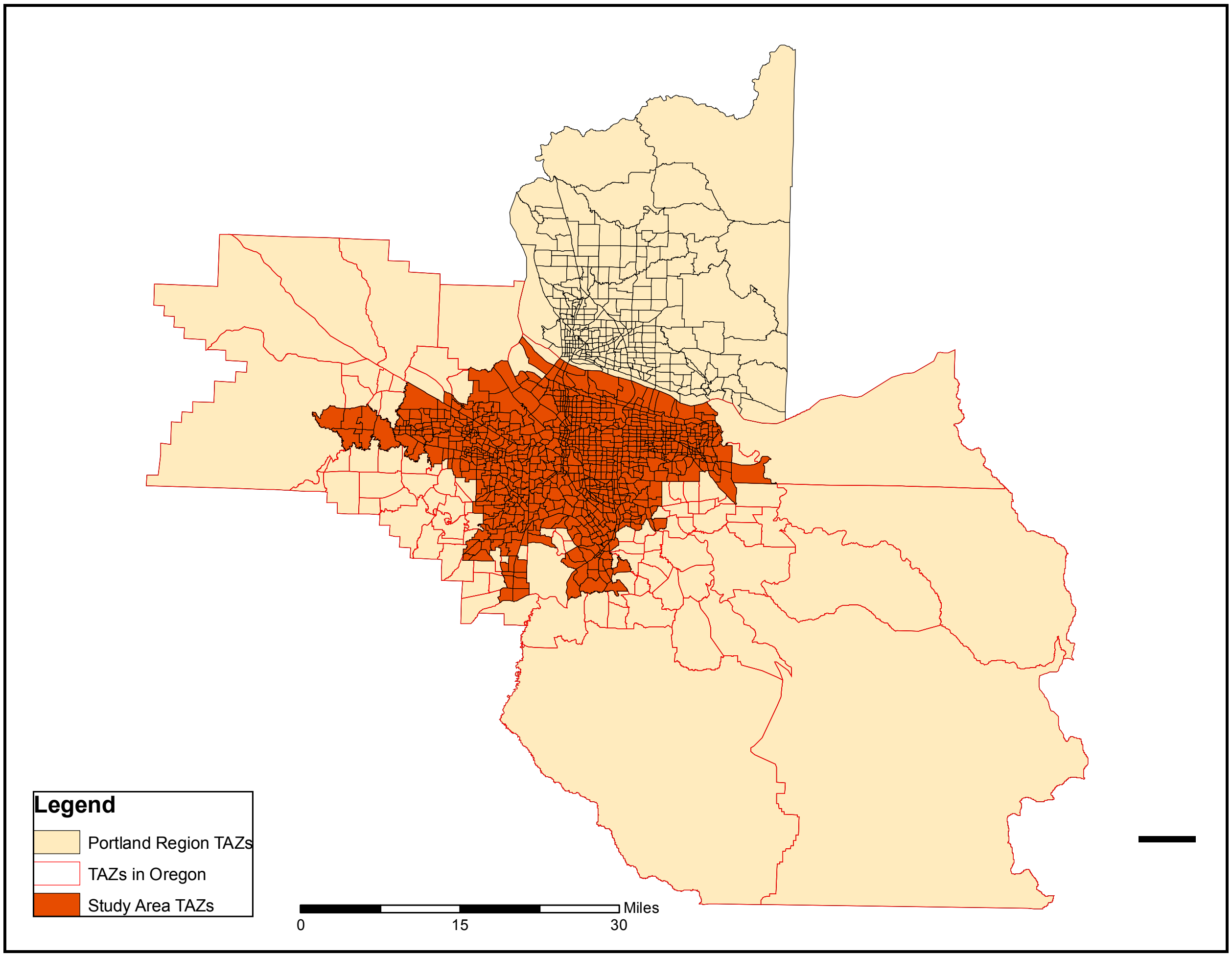


Data obtained from Portland Metro came from a few different locations. Metro publishes the RLIS Lite CD, which contains a broad array of useful information.[9] For the purposes of this analysis, park (land) and bus lines and light rail stops (transit) data were utilized. The earliest publication date of the CD is 1996. Thus, in order to obtain more historic data, a special request was made to Metro. The 1996 bus line shapefile was the oldest available file. The light rail

[9] Categories of data on the CD include: boundaries, Census population, development, environment, land, places, streets, taxlots, transit, and water.

stations 2003 was modified to reflect 1990 conditions based upon the knowledge of Metro staff.[10] The 1990 park layer was obtained separately from Metro. Sidewalk coverage files were also obtained from Portland Metro. These files were created in 2002 and are currently the oldest files available.

Figure 3. Portland TAZs and Study Area Detail.



5.2. <u>DATA PREPARATION WITH A GEOGRAPHIC INFORMATION SYSTEM</u>

5.2.1. Calculation of Development Intensity Factors

Population Density

Population density was measured as persons per gross acre. This analysis uses data from Census 1990 obtained from GeoLytics CD at the block group level (Figure 10 in Appendix).[11] Alternative sources of data were RLIS Lite and CTPP 1990 Part I, Residential. Data from Census 1990 and RLIS Lite were most similar while data from the CTPP varied from the other

[10] Email correspondence with Mark Bosworth at Metro (April 1, 2005).
[11] The field used from this source was "totpop90".

sources. These sources were compared for verification purposes but the Census 1990 data was used in the final analysis.

The data was processed in ArcGIS 9.0. In order to assign population counts from the Census block group shapefile to the TAZs, the Intersect tool in ArcGIS Toolbox was employed. This tool cut the Census block polygons where the TAZ polygons intersected them, in some cases forming new polygons. Each Census block polygon was assigned the information from the TAZ with which it spatially overlaped. The new polygon area was recomputed and population was recalculated as a ratio of the new block area to the old block area. Lastly, the total population was summed for each TAZ.

Some challenges were encountered using the Intersect tool with the Census shapefiles from the GeoLytics CD. Block groups were counted multiple times after intersect, which appeared to be caused by a variation in the Census file. In order to avoid this problem, the Census population attribute table was joined to the RLIS Lite block group shapefile on the common tract/ block group number. The modified population shapefile was then intersected as described above with the TAZ shapefile

Housing Unit Density
Housing unit density was measured as housing units per gross acre. The data for housing unit density was retrieved from CTPP 1990, Part I Residence, which was found on the Census website in tabular format.[12] The scale of resolution is the Census tract (Figure 11 in Appendix). Data was joined to RLIS tract shapefiles by tract number and intersected with the TAZ polygon shapefile as described above.

Employment Density
This factor was measured as jobs per gross acre (Figure 12 in Appendix). Job data (number of workers) was retrieved from CTPP 1990, Part II Work in a tabular format.[13] The same method was used to derive the measure as described above for housing unit density.

Park Density
This factor was measured as parks per gross acre (Figure 13 in Appendix). Park information was obtained from Portland Metro for the year 1990. The data is a polygon layer containing the park type, location, size and other identification information. Twelve different park types were listed in the data dictionary. However, only parks of type "park" were included in this analysis.[14] The Intersect tool was used to count the number of parks per TAZ. The challenge encountered in processing this file was that a single park was frequently composed of many polygons. This required the merging of polygons on the name attribute using the Dissolve tool from the ArcGIS Toolbox. The new polygons were then intersected with the TAZs as described above.

[12] Field name is "U158" and field description is "Total housing units".
[13] Field name is "U204_0101" and field description is "Workers of both sexes, and Total class of worker".
[14] Park types include park, open space, common area of a subdivision or condominium complex, cemetery, golf course, school, pool, tennis courts, fairgrounds/ stadium use, community center, trail/path, and community garden.

5.2.2. Calculation of Motorized Transportation Factors

Roadway

Roadway density is measured as roadway centerline miles per gross acre (Figure 14 in Appendix). Roadway line segments were obtained from 1990 TIGER files. All street types are included in the analysis. For the GIS analysis, the roadway segment polyline file was intersected with the TAZ polygons. The new length of roadway segments was then recalculated. New road length was summed for each TAZ.

Bus Route Density

Bus route density was measured as bus route centerline miles per gross acre and included overlapping routes (Figure 15 in Appendix). The *TriMet Bus System (Routes)* originated from TriMet in 1996. No older digital maps were available. The *Buslines* shapefile is registered to the streets file, enabling horizontal accuracy. Inbound and outbound buses were counted as separate routes for the purposes of this analysis. The bus route polyline shapefile was intersected with the TAZ polygon shapefile. As with the roadway analysis, the new length of the routes was calculated and summed for each TAZ.

Transit Commuting

Transit commuting was measured as percent commuting by public transit mode. Data was obtained from the Census for block groups (Figure 16 in Appendix). As mentioned in population density calculations above, the Census file did not perform as anticipated when intersected with TAZs. It was therefore necessary to join the Census data to the RLIS Lite 1990 Census block group shapefile before the data was intersected with TAZs, as described above. Percent commuting by transit was calculated as the ratio of workers 16 years and older commuting by transit to the total percentage commuting by all modes.[15] Commuting by transit was calculated as the sum of those riding the bus, rail, streetcar or subway.[16]

Proximity to Subway Station

Proximity to subway station was measured as the number of light rail stations within a half-mile of the TAZ center (Figure 17 in Appendix). This analysis used *Light Rail Stations* data from 2003 RLIS Lite transit files. The *LRT_Stop* shapefile is a point file published in 1990 containing the names and locations of stops for the MAX light rail line and central city streetcars. As this file had been updated through 2003, it was necessary to modify the file to reflect the 1990 conditions. Through conversations with Metro staff, it was determined which stops existed in 1990. All other stops were excluded from the analysis. The TAZs were converted to centroids and the light rail stops within a half-mile buffer were counted.

5.2.3. Calculation of Pedestrian and Bicycle Infrastructure Factors

Sidewalk Density

The sidewalk density, measured as sidewalk miles per gross acre, was calculated from the Portland Metro file (Figure 18 in Appendix). The spatial extent of this data includes areas in the urban growth boundary of Portland and the City of Vancouver, Washington. The data was

---

[15] Fields summed for total commuting: DrvAlone, DrvBicyc, DrvBus, DrvCarPo, DrvFerry, DrvMotor, DrvOther, DrvRail, DrvStCar, DrvSubwa, DrvTaxi, and DrvWalk. The number of people working at home was excluded from this calculation (DrvWkHom).

[16] Fields summed for transit commuting: DrvBus, DrvRail, DrvStCar, and DrvSubwa.

published in 2002. This tabular file contains a key field called *LocalID,* which was used to join the table to the streets shapefile.[17] The sidewalk length is represented as a percentage of the total street length. Fields *Lpct* and *Rpct* contain the percentage of street length that is covered by sidewalk for the left and right sides of the street, respectively. Of the 94,095 street segments in the street shapefile, 76,270 matched sidewalk records while 17,825 did not match any sidewalk information. These areas without sidewalk information are located on the periphery of the study area and may have sidewalks that have not yet been mapped. As discussed in data sources above, this constraint limited the extent of the study area.

The GIS approach to the sidewalk analysis involved dividing the sidewalk-street file into two categories. The first group included sidewalks completely within the TAZ boundaries. These sidewalk lengths were summed for each TAZ. The second group of sidewalks, those that were crossed by the TAZ boundary, was intersected with the TAZ shapefile. After intersection, the new sidewalk length was calculated as a ratio of the new street length to old street length multiplied by the percentage sidewalk length (*Lpct* or *Rpct*). All sidewalk lengths were then summed for each TAZ.

Sidewalk Coverage

The sidewalk coverage, measured as sidewalk miles per roadway centerline miles, was calculated in the same manner as sidewalk density described above.

Pedestrian and Bicycle Commuting

Pedestrian and bicycle commuting was measured as the percent commuting by pedestrian or bicycle mode. This calculation used Census 1990 data and was derived from the same data as that for transit commuting (Figure 19 in Appendix). The percentage commuting by pedestrian or bicycle modes was calculated as the ratio of pedestrian and bicycle modes to the total number of workers 16 years and older commuting.[18]

5.3. CALCULATION OF BEI FOR EACH TRAFFIC ANALYSIS ZONE

Three different statistical approaches are used to classify the TAZs into categories for the purposes of comparison with the PEF (naïve ranking, cluster, and principal components analysis, PCA). The BEI was partitioned into three groups, urban, suburban, exurban. The classification is somewhat arbitrary, as it does not follow pre-established measurements for the different built environments. Furthermore, it relies upon the assumption that the range from urban to exurban is fully represented within the study area. However, this nominal classification facilitates understanding of the index and can be more meaningful than the numerical rankings.

The first method of classification is the naïve ranking approach. Each of the eleven factors in the analysis is partitioned into six groups based on the percentile that it falls into. For example, the range of values for population density is divided into six percentiles, the top sixth are assigned the number six and the lowest ranking sixth are assigned zero. In this way each factor for every

[17] The RLIS Lite street shapefile from 2003 was used as the base shapefile for sidewalks. This was necessary because the sidewalk data was in a tabular format. In order to perform spatial queries, the sidewalk data table was joined to the street shapefile on a shared Local ID number.

[18] Fields summed for pedestrian and bicycle commuting: DrvBicyc and DrvWalk.

TAZ receives a ranking comparing it to all the other values across the TAZs. Proximity to subway is more heavily weighted than other factors and receives simply either the value ten if there are one or more stops within a half-mile buffer of the TAZ centroid and a zero otherwise.

The rankings for each TAZ are averaged across the factors that comprise each of the three domains (development intensity, motorized transportation, and pedestrian and bicycle infrastructure). The BEI is calculated by adding up the scores, which are weighted by the elasticity of demand, as determined by Cervero and Kockelman (1997).[19] According to this San Francisco Bay Area study, average elasticity of demand for transportation and motorized transportation was close to 0.099 while the average for pedestrian and bicycle modes was 0.16, or roughly 1.6 times more responsive to walking quality factors.[20] The rank scores are then classified into three groups based upon a natural breaks method. This algorithm seeks to classify data on inherent breaks so that similar data is grouped together while maximizing differences between classifications. It seeks to maximize the goodness of variance fit by minimizing the squared deviations of group means.[21]

The second approach is the non-hierarchical cluster analysis using k-means methodology. This exclusive clustering approach classifies data into k clusters. In this case, the k is set to three clusters. The method assigns data points to the closest cluster centroid while optimizing the distance between clusters. The process of moving cluster centroids in the dataset is repeated until their location is stable. This method is sensitive to the initial placement of the centroids but repetitive analysis can address this issue to some degree. Additionally, this analysis assumes three clusters best describe the data, when in fact the data may actually have more or less clusters.

The third approach is principal components analysis (PCA). This approach, based on linear algebra, extracts the primary variables that best capture the data from a large set of variables. It seeks to re-express the data set in a new basis, reducing the noise and allowing relationships to be viewed more simply. It assumes linearity, Gaussian distribution of the data (the mean and standard deviation sufficiently describe the data), orthogonal principal components, and assumes directions with the greatest variance are the most important components. First, a normalized direction of greatest variation is selected. Then a second direction of maximal variation perpendicular to the first direction is selected (satisfying the orthonormal assumption). The outcome of this analysis is a single formula used to predict the BEI score for each TAZ. As in the naïve approach, the natural breaks method is then used to classify the PCA scores into three categories.

---

[19] Formula for calculating the BEI from the average domain rank scores.
BEI = 0.099 (domain 1) + 0.099 (domain 2) + 0.16 (domain 3)

[20] Elasticity of travel demand for measures of built environment for non-personal vehicle trips is listed for non-work, personal business, and work trips separately. Here we use the average of these trip types for the weights associated with each domain. Density = 0.099 [(0.084 + 0.113)/ 2]. Design (walking quality factor) = [(0.183 + 0.174 + 0.119)/3]

[21] Esri Support Center. What is the Jenks optimization method? ESRI. <http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=26442> March 25, 2005.

6. RESULTS

This section describes the general outcomes of the three approaches (naïve ranking, cluster, and PCA), compares them with each other, and compares them with the PEF. Of the original 1,247 TAZs, 873 were analyzed (see discussion in *Data Sources* for an explanation why certain TAZs were removed from the analysis). The outputs were generally as expected and have been summarized in Table 2.

The average factor values for the three classifications (urban, suburban, and exurban) for each approach are listed in Table 7 (Appendix). The averages increase as expected from exurban to urban. Some values vary considerably across the range, such as population density, employment density, and percent commuting, while other values are relatively similar between the three classifications such as roadway, bus routes, sidewalk density, and especially park density.

Table 2. Factor Statistics.

Factor	Measure	Mean	Standard Deviation	Min	Max
Development Intensity Factors					
Population density	Population/gross acre	5.53	4.06	0.01	27.57
Housing unit density	Housing units/gross acre	1.96	2.08	0.00	20.82
Employment density	Jobs/gross acre	3.62	12.83	0.00	204.19
Park density	Parks/gross acre	0.01	0.01	0.00	0.09
Motorized Infrastructure Factors					
Roadway density	Roadway centerline miles/gross acre	0.02	0.01	0.00	0.07
Bus route density	Bus route centerline miles/gross acre (includes overlapping routes)	0.02	0.04	0.00	0.61
Transit commuting	Percent commuting by public transit mode	0.07	0.06	0.00	0.47
Proximity to subway station	Metro rail station within 0.5 miles or 1 mile of the CAZ center	0.04	0.20	0.00	1.00
Ped/ Bike Infrastructure Factors					
Sidewalk density	Sidewalk miles/gross acre	0.03	0.03	0.00	0.13
Sidewalk coverage	Sidewalk miles/roadway centerline miles	0.93	0.58	0.00	2.00
Ped and bicycle commuting	Percent commuting by pedestrian or bicycle mode	0.05	0.07	0.00	0.55

The naïve ranking method split the TAZs into six percentile groups (Table 3). Each factor value was reassigned a rank score based upon the percentile range it fell within. Park density presented a challenge because the percentile distribution was such that both the first and second percentile divisions had zero values. This was addressed by artificially deflating the park scores so that all raw values of zero were assigned a rank value of zero. Proximity to subway station was also handled differently than the other factors. These scores were equalized such that all

TAZs containing one or more light rail stations within a half-mile of the TAZ center were assigned a rank value of ten (otherwise assigned zero). This created more than a doubling of the variable *proximity to subway station* in the motorized infrastructure domain. All scores were averaged for each domain and the BEI score was calculated based on the weighted formula (see *Methods*).

Three different measures of dissimilarity were performed in order to test which approach was most appropriate (Euclidian distance, the Euclidian distance squared, and the Canberra distance). Euclidian cluster analysis is presented here and subsequently compared with the other two approaches, naïve and PCA, since it most closely correlated with them based on the Pearson correlation coefficient for the raw scores.

The results of the PCA yield a formula for predicting the BEI values for each TAZ. This formula explains 45.7% of the variation observed in the dataset. The score for each factor indicates its relative importance in determining the BEI (Table 4). Sidewalk density, roadway density, and housing unit density were the three most important variables in this dataset.

Table 3. Naive Ranking Scoring (Percentiles).

	Percentiles					
	16.67%	**33.33%**	**50.00%**	**66.67%**	**83.33%**	**100.00%**
Rank Score	**0**	**1**	**2**	**3**	**4**	**5**
Factors						
Development Intensity Factors						
Population density	1.278	3.249	5.076	6.848	9.206	27.567
Housing unit density	0.212	0.618	1.510	2.331	3.583	20.817
Employment density	0.180	0.640	1.256	2.041	3.709	204.194
Park density	(all values below 0.030 assigned 0)			0.006	0.011	0.091
Motorized Infrastructure Factors						
Roadway density	0.010	0.017	0.024	0.029	0.039	0.072
Bus route density	0.000	0.004	0.008	0.013	0.022	0.613
Transit commuting	0.023	0.037	0.051	0.073	0.112	0.467
Proximity to subway station	(either 0 or 10; 0 for no stations, 10 for one more stations within 1/2 mile buffer)					
Ped/ Bike Infrastructure Factors						
Sidewalk density	0.005	0.012	0.021	0.035	0.052	0.128
Sidewalk coverage	0.295	0.572	0.886	1.211	1.628	2.000
Ped and bicycle commuting	0.012	0.019	0.028	0.039	0.064	0.552

Values listed in Table 3 form the upper bound of the range.

Table 4. Principal Components Analysis Formula.

Factors	Relative Importance*	PCA Score
Development Intensity Factors		
Population density	5	0.3330
Housing unit density	3	0.3657
Employment density	6	0.2929
Park density	11	0.1444
Motorized Infrastructure Factors		
Roadway density	2	0.3705
Bus route density	9	0.2430
Transit commuting	4	0.3358
Proximity to subway station	10	0.2066
Ped/ Bike Infrastructure Factors		
Sidewalk density	1	0.3807
Sidewalk coverage	7	0.2869
Ped and bicycle commuting	8	0.2641

* Importance from 1 to 11, where 1 is most important.

6.1. COMPARISON OF APPROACHES

The naïve ranking approach varied most from the two other approaches in the number of TAZs classified into each of the three categories (urban, suburban, and exurban). The naïve approach classified roughly 27% as urban, 43% as suburban, and 30% as exurban (Table 7 in Appendix). On the other hand, both cluster and PCA classified approximately 14% as urban, 37% as suburban, and 49% as exurban. The naïve approach groups more TAZs towards the middle of the distribution (suburban), and PCA and cluster skew values towards the end of the distribution (exurban). See Figure 4 to Figure 6 below for classified scores.

The comparison of the raw scores for each of the three approaches using the Pearson correlation coefficient indicates that naïve ranking and PCA scores are more closely correlated (0.89*),[22] although there is a high degree of correlation between all scores (Table 5). PCA and cluster approaches have the lowest Pearson correlation coefficient of 0.84* and naïve and cluster have a coefficient of 0.86*.

The outcome of the comparison of the final scores using the kappa statistic and percent agreement was slightly different.[23] Here, PCA and cluster approaches had the highest percent agreement (92.9%) and the highest kappa statistic (0.88*).[24] Naïve and cluster had 69.6%

---

[22] Indicates scores are significant at 99% confidence level (p = 0.01).
[23] As indicated in the *Methods* section, the final score for each approach is 1, 2, or 3, representing exurban, suburban, and urban respectively.
[24] A kappa statistic of 1 indicates perfect agreement and 0.8 is generally accepted as almost perfect agreement (Landis and Koch, 1977).

agreement and a kappa of 0.54* while naïve and PCA had the lowest agreement at 65.9% and a kappa of 0.48*. The comparison of the raw scores indicated a slightly better correlation between naïve and PCA, which was opposite to the outcome with the final scores, where PCA and cluster were better correlated.

Each of the three measures of agreement used to verify the comparison indicated a strong degree of correlation, confirming the integrity of the analysis. This indicates that one may use the BEI with a degree of confidence. Additionally, it is suggested that the implementation of the index does not require the use of all three approaches but can employ the approach that best fits the requirements and capabilities of the situation.

Table 5. Comparison of Approaches.

Pearson correlation (raw scores)		
	Cluster	**PCA**
Cluster	1.00	
PCA	0.84	1.00
Naïve	0.86	0.89
% Agreement (classified scores)		
	Cluster	**PCA**
Cluster	100.0%	
PCA	92.9%	100.0%
Naïve	69.6%	65.9%
Kappa (classified scores)		
	Cluster	**PCA**
Cluster	1.00	
PCA	0.88	1.00
Naïve	0.54	0.48

All scores significant at a 99% level of confidence.

Figure 4. Naive Ranking Classification of Study Area.

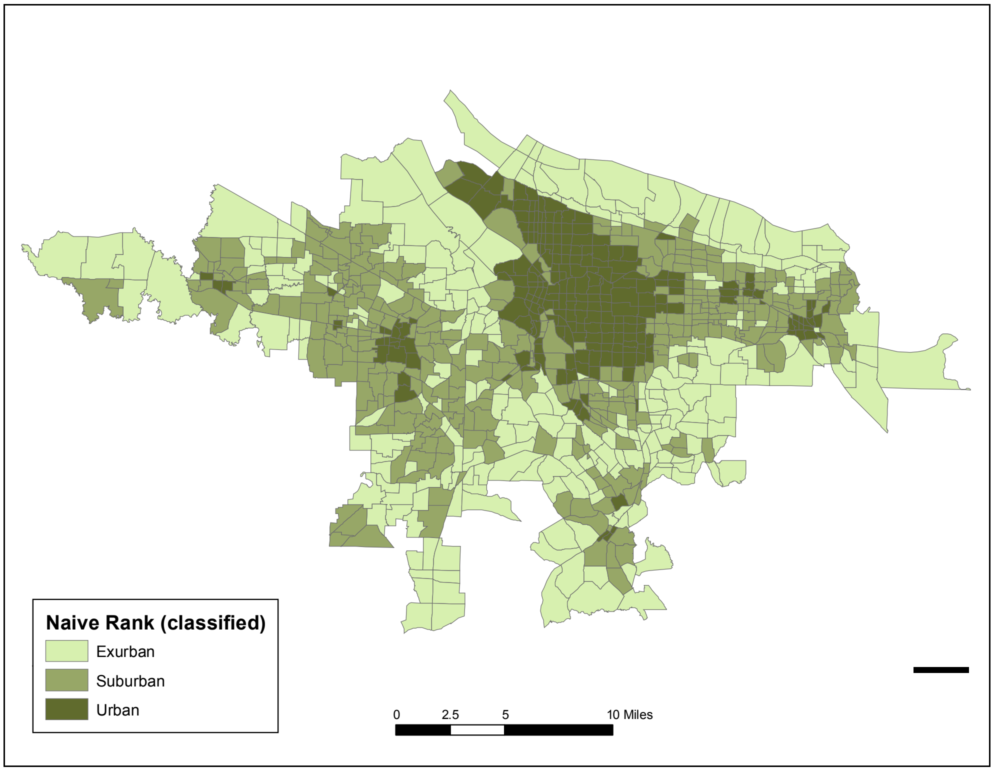


Figure 5. Cluster Analysis Classification of Study Area.

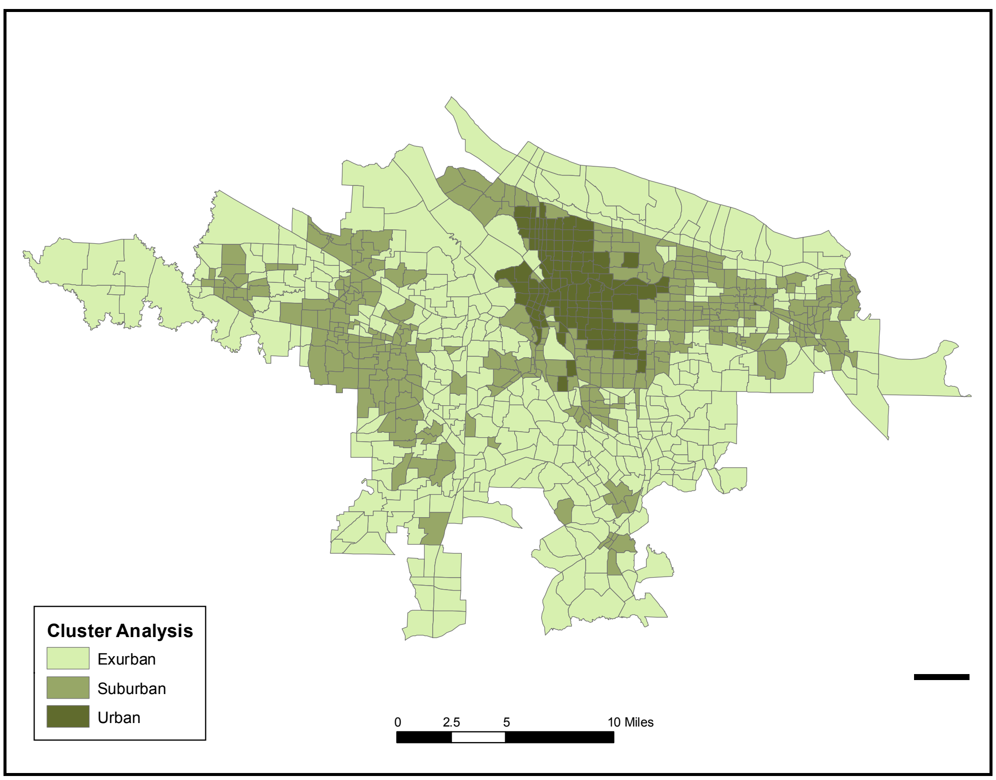
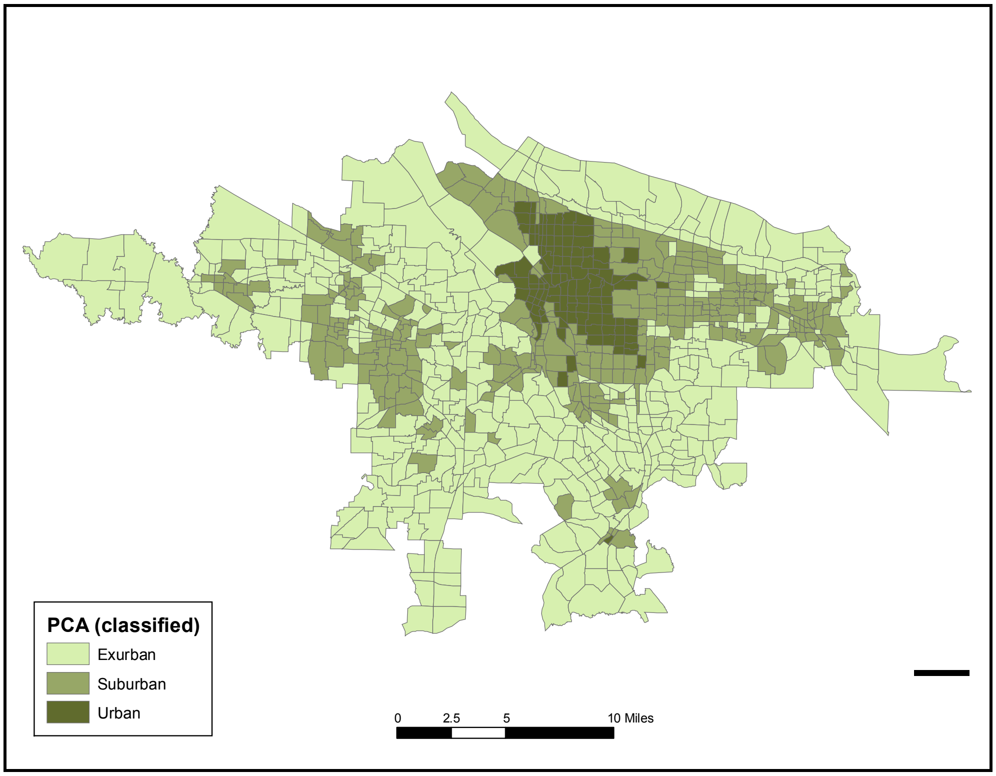


Figure 6. PCA Classification of Study Area.



6.2.COMPARISON WITH THE PEF

The PEF was first compared with the raw scores for the three approaches (Table 6). The Pearson correlation coefficient was higher for both naïve (0.71*) and PCA (0.71*) and lower for cluster (0.66*).[25] Next the PEF was classified into three groups using natural breaks in order to compare it with the final scores for each approach (this is the same technique used to classify the naïve and PCA raw scores into final scores.

Table 6. Comparison with PEF.

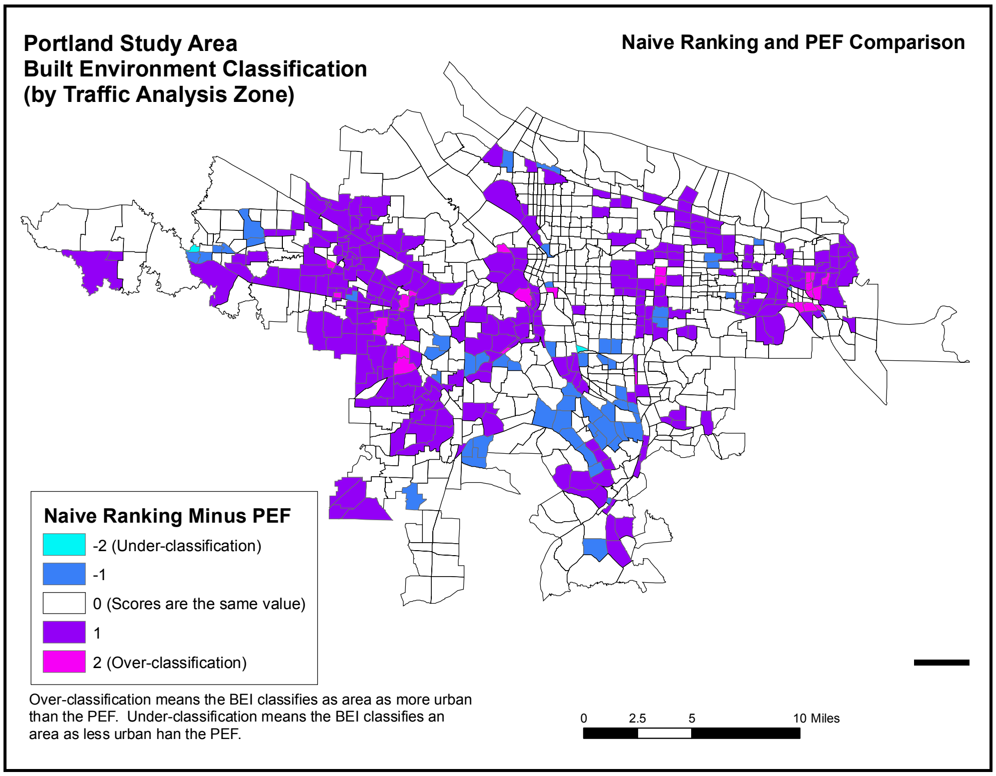
Approach	Pearson (raw scores)	% Agreement (final scores)	Kappa (final scores)
PCA	0.71	67.2%	0.46
Naïve	0.71	58.8%	0.38
Cluster	0.66	64.2%	0.42

All scores significant at a 99% level of confidence.

[25] It should be noted that Pearson's coefficient should be used with continuous variables. In this case, the low score for cluster may be partly explained by the fact that it is not a continuous variable.

PCA had a slightly higher percent agreement with the PEF (67.2%) and a kappa statistic of 0.46*. Both cluster and naive agreed less with the PEF (64.2% and 58.8% respectively) and had lower kappa statistics (0.42* and 0.38* respectively). The low kappa statistic indicates that the BEI and PEF do not capture the same concepts (see Figure 7 to Figure 9). Another consideration is that the pedestrian and bicycle infrastructure factors, the third domain, may more closely represent the factors contained in the PEF. The naïve ranking score averages for each of the three domains were compared with the PEF. However, the Pearson correlation coefficient for each of these domains was even lower than the correlation with the entire indices.[26] The synergy between the various factors within the BEI therefore adds value to the overall score.

Figure 7. Comparison of Naive Ranking with PEF Scores.



[26] Pearson coefficients for rank score averages for the domains are as follows: development intensity factors (0.66), motorized infrastructure factors (0.65), and pedestrian and bicycle infrastructure factors (0.58). All values are significant at the p = 0.01 level.

Figure 8. Comparison of Cluster Analysis with PEF Scores.

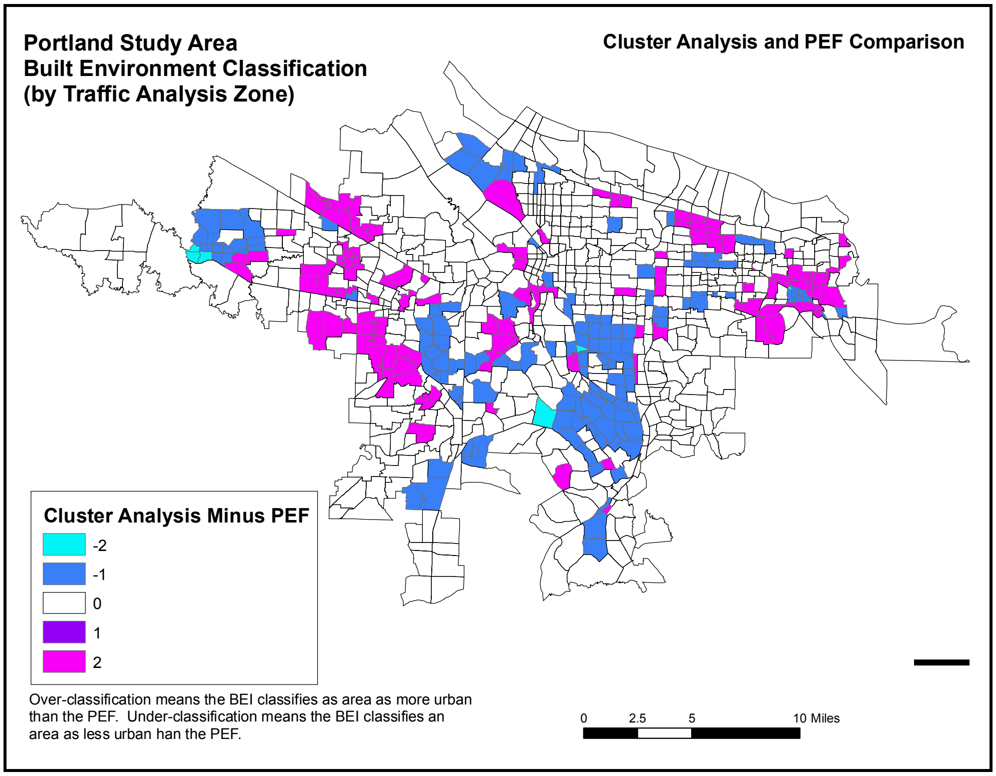
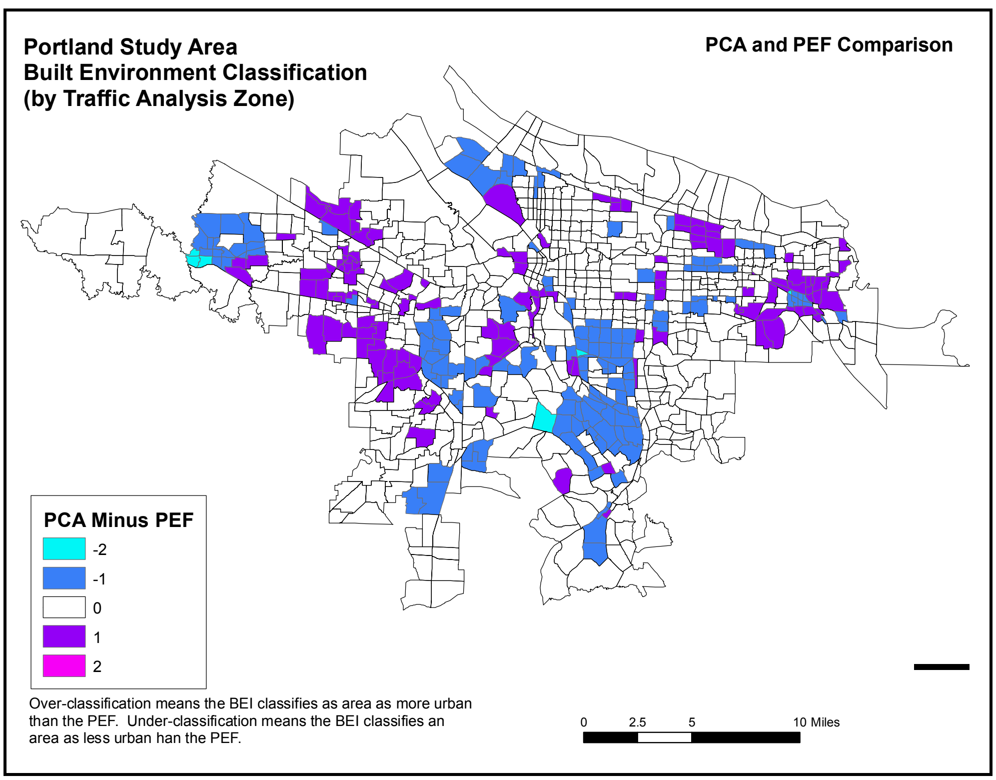


Figure 9. Comparison of PCA with PEF Scores.



7. DISCUSSION

The initial purpose of this project was to compare the BEI with the PEF in order to see how the two methods varied. Indeed, they vary not only in the numbers and types of factors that compose them, the nature of the method (desktop versus field survey), the approaches (subjective Delphi-like process versus objective rank/factor/cluster analysis), but also in the outcome. While they share connectivity and accessibility concepts, they measure these concepts in distinct ways. The low kappa statistic and comparison scores indicate that the BEI is not a replacement for the PEF.

The BEI may be preferable to field survey approaches for several reasons. The primary considerations are the large savings in time and costs. The BEI utilizes previously collected data and does not require considerable staff resources. It is can be implemented relatively easily using GIS and a basic statistics package (Excel, SPSS, STATA). A secondary advantage of the BEI is that it yields an objective analysis independent of the perspective of the researcher. For indices such as the PEF, issues of consistency between field survey scores complicate inter-regional comparisons where different individuals create the scores. On the other hand, the BEI allows for broad-scale, inter-regional analysis.

There are several possible ways in which the BEI could be improved. One criticism of the approach is that is does not directly account for mixed land use. Many studies have shown that this is a critical variable in alternative transportation models (Cervero and Kockelman, 1997; Hess and Ong, 2001; LCOG, 2003). Additionally, the types of land use mixes appear to play an important role in pedestrian/ bicycle activity (Ewing and Cervero, 2001; Krizek, 2003). The BEI does not contain a land use mix variable per se, although it does include housing unit density and job density. Including a measure of the entropy, dissimilarity index, or other calculation of mixed use could strengthen the index.

Another way in which the BEI might be modified is to change the units of analysis. The current unit of analysis, the TAZ, is a rather large and heterogeneous unit. Krizek (2003) addressed this challenge for the purposes of measuring neighborhood accessibility by standardizing and reducing of the size of the units to a 150-meter grid. Such an approach, if applied to the BEI, could provide a more manageable scale of analysis. However, the added requirements of using such a grid scale may out-way positive benefits by slowing analysis time or requiring dis-aggregation of data or new data sources (e.g. parcel-based layers).

A third consideration for revision of the BEI is the park factor measurement unit. The current approach of measuring parks per gross acre gives points to TAZs with many small parks but may not properly represent TAZs with larger, but fewer parks. For example, if a TAZ has one large park, it will score lower than a TAZ with a few small parks. Do multiple, smaller parks present more opportunity for physical activity or encourage more non-auto trips than larger parks? One might recommend modifying the measure to account for park acres per gross acre.

The general trend in collecting and combining data for built environment indices seems to indicate a shift from the field survey approach to the GIS approach. Much of this is a result of the changing technology from the early 1990s to the twenty-first century. Spatial data for the region was not as accessible, and in some cases, did not exist. Now much more data is available electronically, allowing the scope and scale of the analysis to expand. It is likely that built environment indices will continue to be refined so that they increasingly capture the varied factors influencing active transportation.

8. CONCLUSIONS

Despite the differences between the PEF and BEI and the low correlation between the two indices, the BEI should still be advanced as a valuable index. The comparison of the three approaches in classifying the GIS analysis revealed a strong degree of internal consistency between the approaches. Based on the literature reviewed, the BEI is a solid index representing the range of land use variables considered important in non-auto modes of travel. It contains density measures, a simple representation of housing/ jobs balance and land use mix (housing, employment, and park density), captures ideas of accessibility and connectivity (sidewalk measures, roadway measures, and percent commuting by walk/ bicycle modes), as well as mobility (bus lines, light rail stations, percent commuting by transit).

There are several caveats to this study that should be noted. First, this study is a single case study. It is restricted to a comparison with the PEF for only one geographic region. Therefore, there are limitations to the generalizations that can be drawn from this analysis. Second, while the goal was to complete the analysis with U.S. Population Census data, other sources had to be used. Bus routes, light rail stations, sidewalk, and park data were not available through the Census. While it is fortunate that this data is available from Metro, not all municipalities will have this information (especially sidewalks), which will limit implementation of the BEI. Additionally, because this study required historic data, it was more challenging to find the files. Older data is not easily accessible from online sources and historic data may not even exist, as was the case for the bus routes (1996) and for sidewalks (2002).

The BEI provides a solid entry point for researchers studying the built environment. As more resources become available, the BEI can be tailored to the particular needs of the study. Focused research questions can guide modifications to the index. Ultimately, the validity of the index could be tested with field surveys as was done by Ulmer and Hoel (2003) to confirm the findings of their index. The BEI can be utilized in travel behavior models (mode choice, trip generation, etc.) as was done with the PEF and PFI in order to increase the predictive validity of the models. It provides communities with a tool to consider land use and transportation policy decisions as it allows them to quantify the built environment, assess strengths and shortcomings, and target areas for infrastructure improvements and policy changes, as demonstrated in Montgomery County, Maryland (Replogle, 1990). New possibilities and applications of indices continue to be found. They can guide individual, commercial, and retail real estate decisions and funding of government programs, such as location efficient mortgages (Ulmer and Hoel, 2003).

One of the greatest strengths of the index is that it is easily accessible to many planning departments and its implementation is straightforward. However, because it does not contain the type of detailed analytic input that is required by a PEF -type index, it lacks the resolution and on-the-ground verification that other approaches might provide. Modifying the index may allow for some gains in accuracy at the neighborhood level but the impact of modifications on the time and ease of implementation of the index should be understood. Increasing the time cost of acquiring data or the degree of sophistication of the GIS analysis may hamper the use of the BEI. It is suggested then that the BEI should be utilized as an overview tool to acquire a first-glance of built environment factors influencing active transportation. To supplement this review, a more fine-grained study of particular neighborhoods would be recommended.

9. REFERENCES

Banerjee, T. and Baer, W. C. (1984). *Beyond the neighborhood unit: residential environments and public policy*. New York: Plenum Press.

Blommaert, M., Borms, J., and Hebbelinck, M. (1981). *Relative influences of urbanization affecting motor play activities in comparison with microsocial and somatic characteristics*. Proceedings of the 24th World Congress on Health, Physical Education, and Recreation, Manila, Philippines, ICHPER, 25–36.

Cambridge Systematics, Inc. and Barton Aschman Associates. (1994). *Short-term travel model improvements.* Travel Model Improvement Program, U.S. Department of Transportation.

Cambridge Systematics, Inc., Parsons, Brinckerhoff, Quade, and Douglas, Inc., S.H.Putman Associates, Inc. (1992). *Making the land use, transportation, air quality connection, Volume 4, Model modifications*. Portland, Oregon: 1000 Friends of Oregon.

Cervero, R. and Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity and design. *Transportation Research D,* 2(3), 199-219.

DKS Associates. (2002). *Model update report- Sacramento regional travel demand model, Version 2001* (SACMET 01). Sacramento Area Council of Governments.

Eash, R. (1997). *Incorporating urban design variables in metropolitan planning organizations' travel demand models.* Urban Design, Telecommunication and Travel Forecasting Conference: Summary, Recommendations and Compendium of Papers, Texas A&M University.

Ewing, R. and Cervero, R. (2001). Travel and the built environment- Synthesis. *Transportation Research Record*, 1780, 87-113.

Frank, L.D., Sallis, J.F., Saelens, B.E., Leary, L., Cain, K., Conway, T.L., and Hess, P.M. (2004). A walkability index and its application to the trans-disciplinary neighborhood quality of life study. *Journal of Planning Education and Research*, (under review).

Greenwald, M.J. and Boarnet, M.G. (2001). The built environment as a determinant of walking behavior: Analyzing non-work pedestrian travel in Portland, Oregon. *Transportation Research Record,* 1780, 33-42.

Handy, S.L., Boarnet, M.G., Ewing, R. and Killingsworth, R.E. (2002). How the built environment affects physical activity -Views from planning. *American Journal of Preventive Medicine,* 23(2S), 64-73.

Handy, S., Cao, X., Buehler, T.J., and Mokhartian, P. (2005). *The link between the built environment and travel behavior: Correlation or causality?* Transportation Research Board 2005 Annual Meeting.

Hess, D.B. and Ong, P.M. (2001). Traditional neighborhoods and auto ownership. *Transportation Research Record*, 1805, 35-44.

Kitamura, R., Laidet, L., and Mokhartian, P. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, 24, 125-158.

Krizek, K.J. (2003). Operationalizing neighborhood accessibility for land use–travel behavior research and regional modeling. *Journal of Planning Education and Research*, 22 (3), 270-287.

Landis, J. R. and Koch, G. G. (1977). The measurements of observer agreement for categorical data. *Biometrics,* 33, 159-174.

Lane Council of Governments. (2003). *Statistical analysis of urban design variables and their use in travel demand models.* Performance Measures Subcommittee of the Oregon Modeling Steering Committee, Oregon Department of Transportation.

Lynch, K. (1981). *Good City Form*. Cambridge: MIT Press.

Parsons, Brinkerhoff, Quade, and Douglas, Inc. (2000). *Data collection and modeling requirements for assessing transportation impacts of micro-scale design* (DTFH61-95-C-00168). Washington, DC: Federal Highway Administration.

Parsons, Brinckerhoff, Quade, and Douglas, Inc., Cambridge Systematics, Inc., and Calthorpe Associates. (1993). *Making the land use, transportation, air quality connection, Volume 4A, The pedestrian environment*. Portland, Oregon: 1000 Friends of Oregon.

Sallis, J.F., Nader, P.R., Broyles, S.L., Berry, C.C., Elder, J.P., McKenzie, T.L., and Nelson, J.A. (1993). Correlates of physical activity at home in Mexican-American and Anglo-American preschool children. *Health Psychology*,12, 390–398.

Sallis, J.F., Frank, L.D., Saelens B.E., and Kraft, M.K. (2004). Active transportation and physical activity: Opportunities for collaboration on transportation and public health research. *Transportation Research Part A*, 38, 249–268.

Schlossberg, M. and Brown, N. (2004). *Comparing transit oriented developments based on walkability indicators*. Transportation Research Board 2004 Annual Meeting.

Replogle, M. (1990). Computer transportation models for land use regulation and master planning in Montgomery County, Maryland. *Transportation Research Record,* 1262, 91-100.

Ulmer, J. and Hoel, L.A. (2003). *Evaluating the accessibility of residential areas for bicycling and walking using GIS* (UVACTS-5-14-64). University of Virginia, Center for Transportation Studies.

10. APPENDIX

Figure 10. Persons Per TAZ (Raw Data).

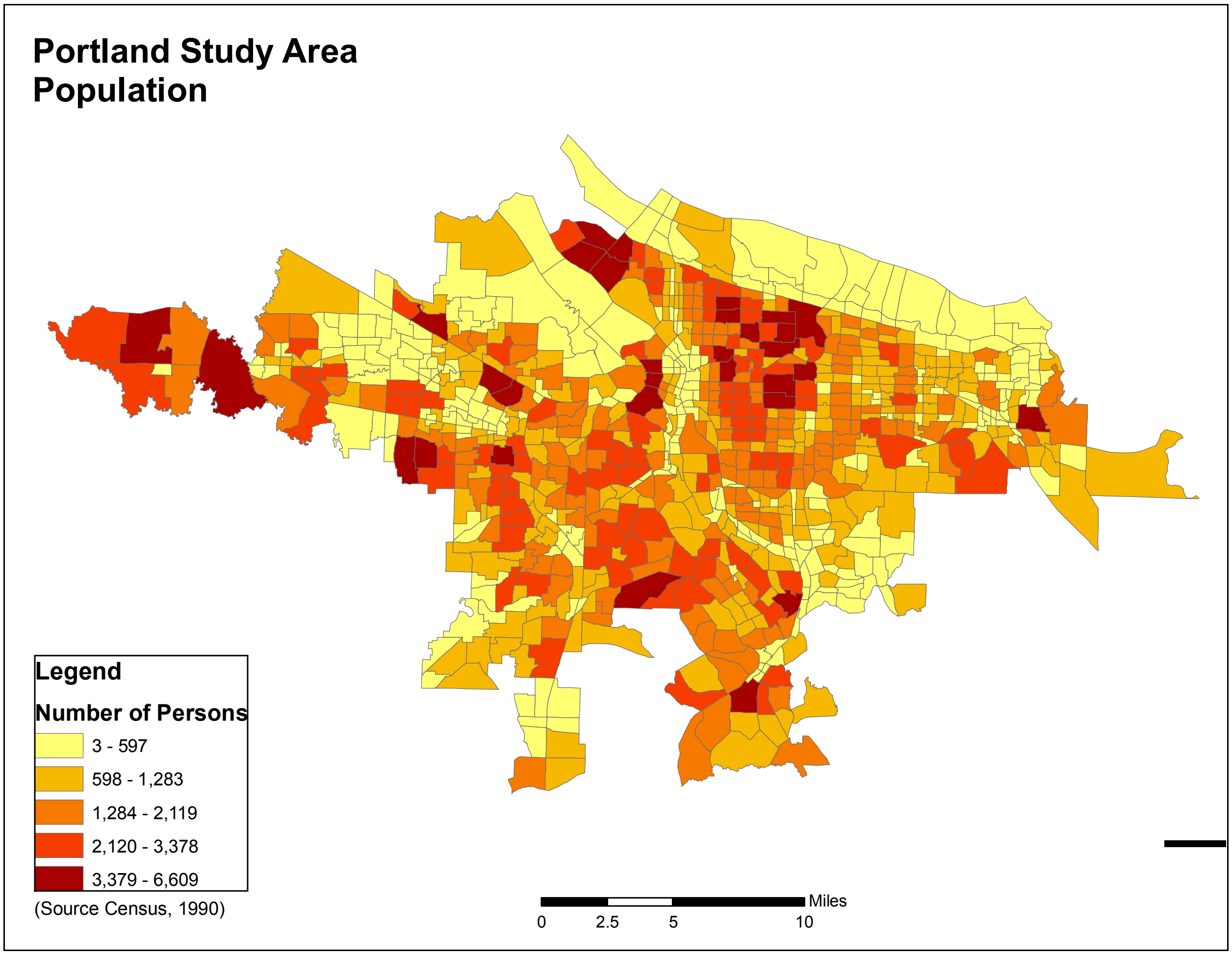


Figure 11. Housing Units Per TAZ (Raw Data).

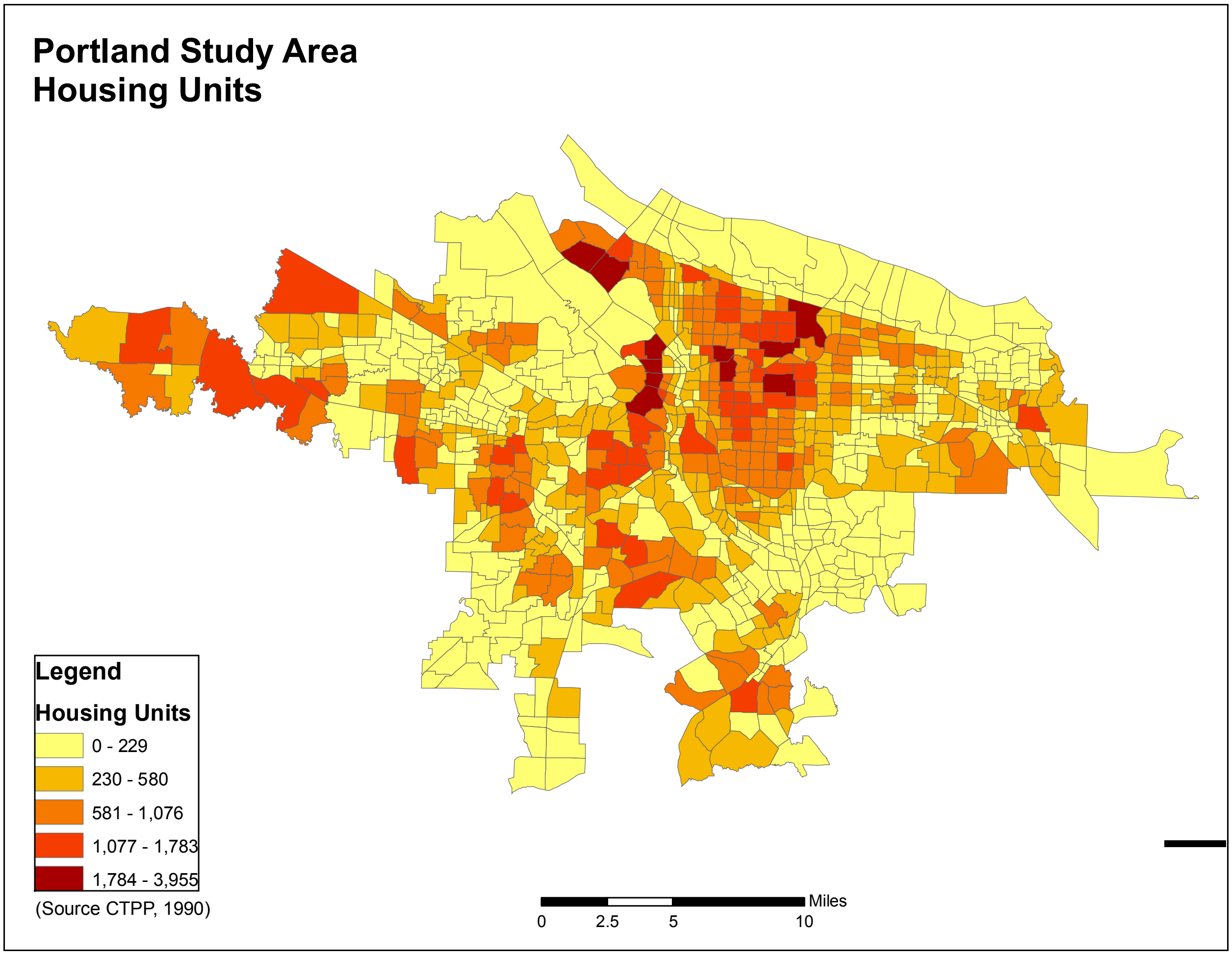


Figure 12. Number of Jobs Per TAZ (Raw Data).

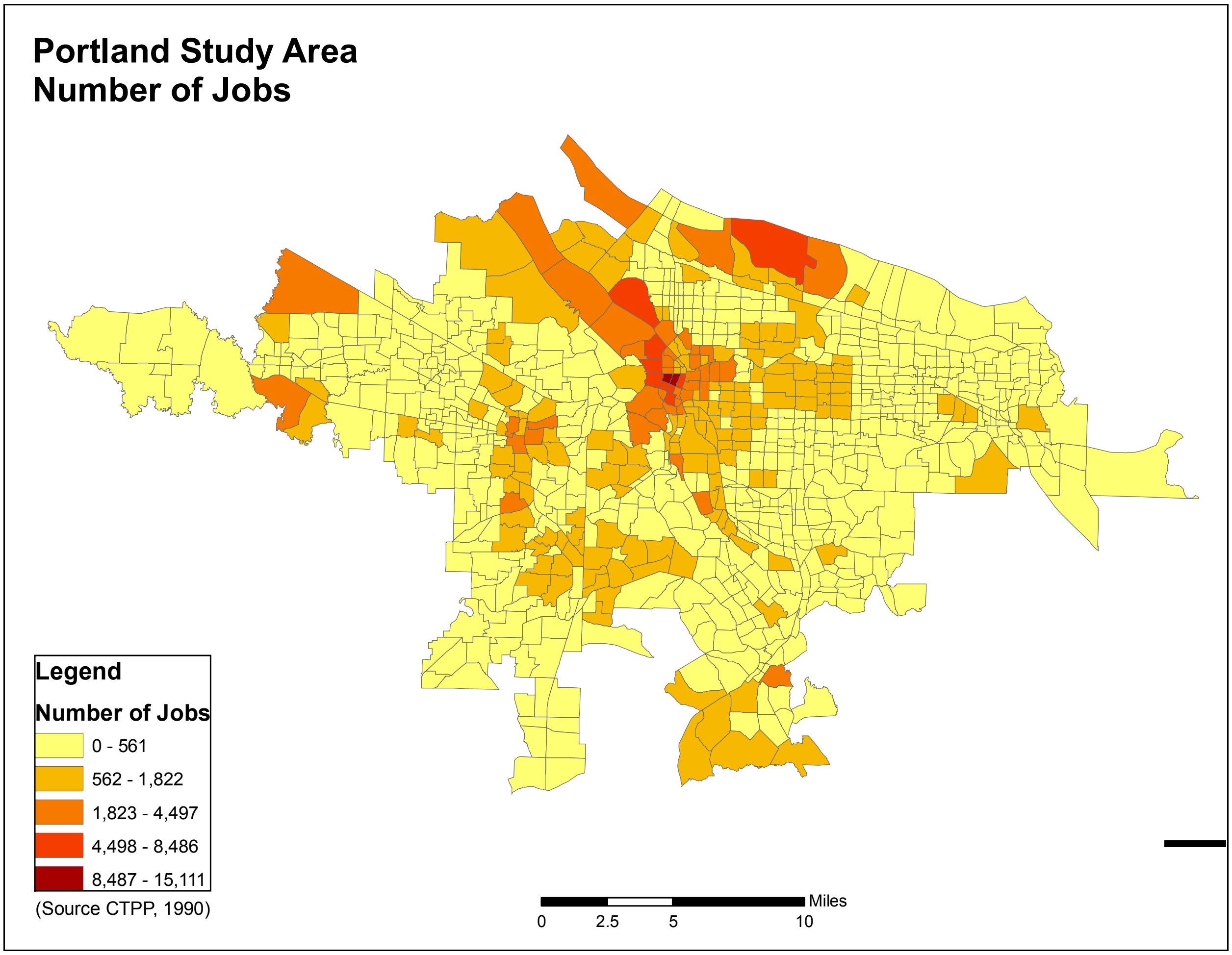


Figure 13. Parks (Raw Data).

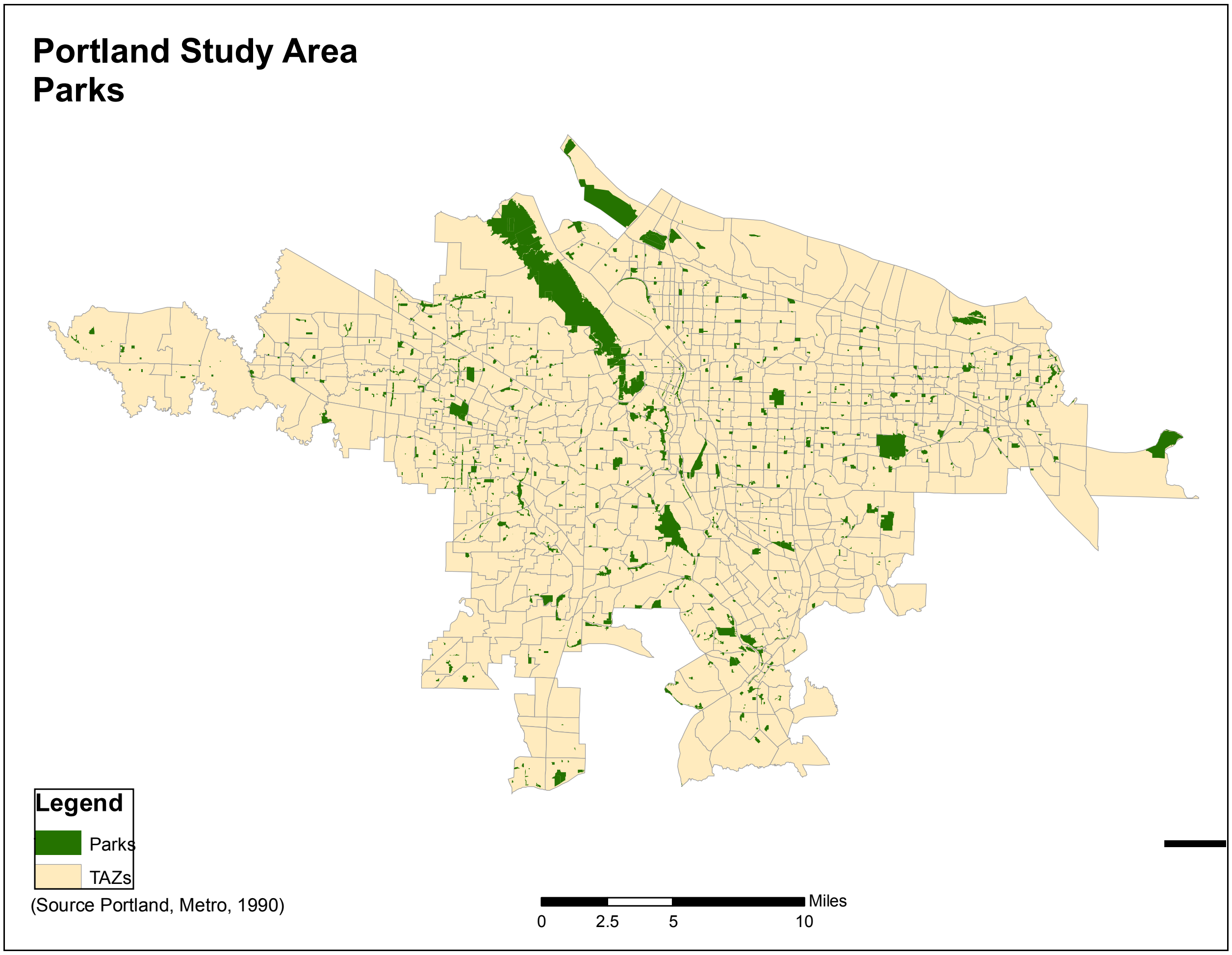


Figure 14. Roads (Raw Data).

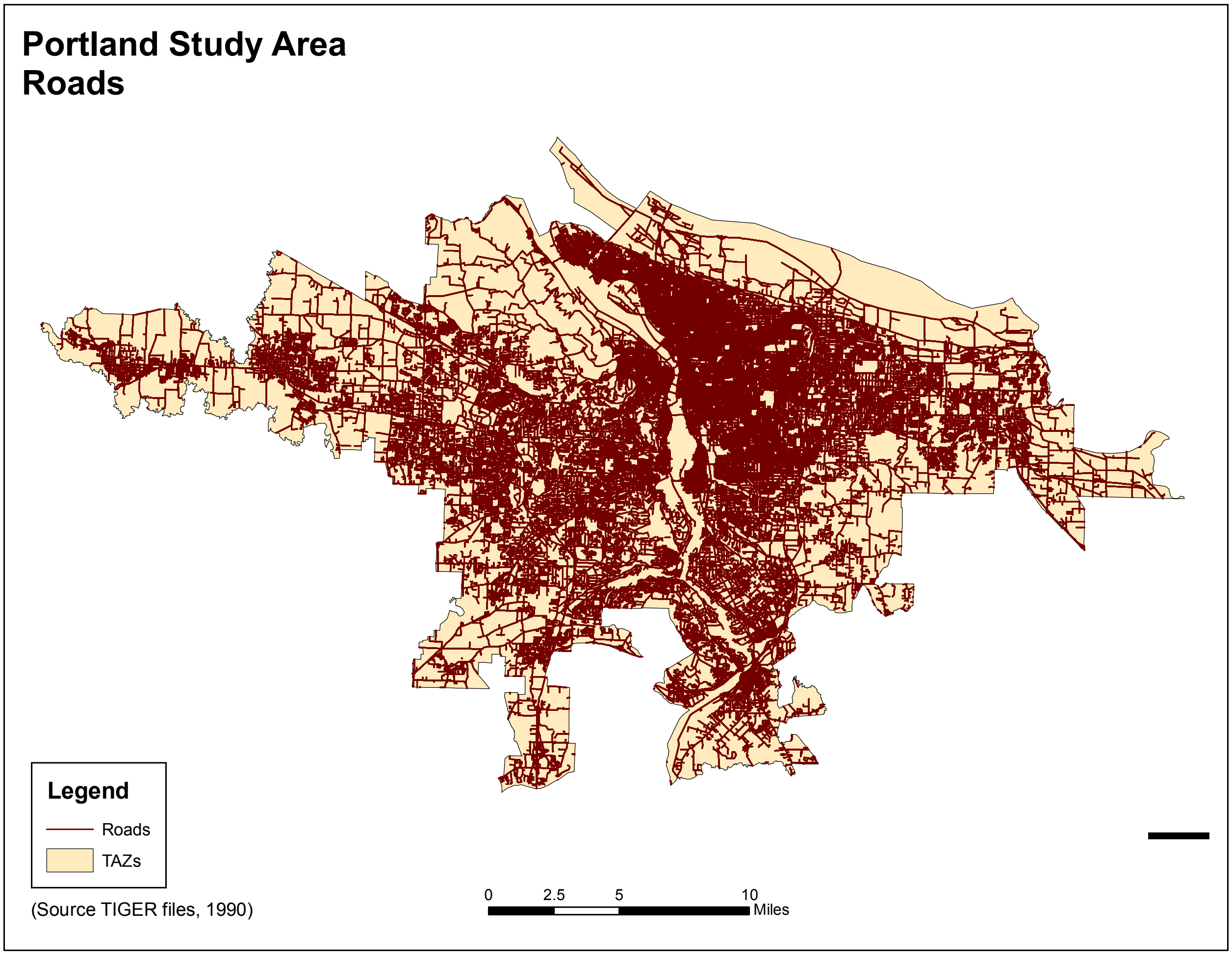


Figure 15. Bus Routes (Raw Data).

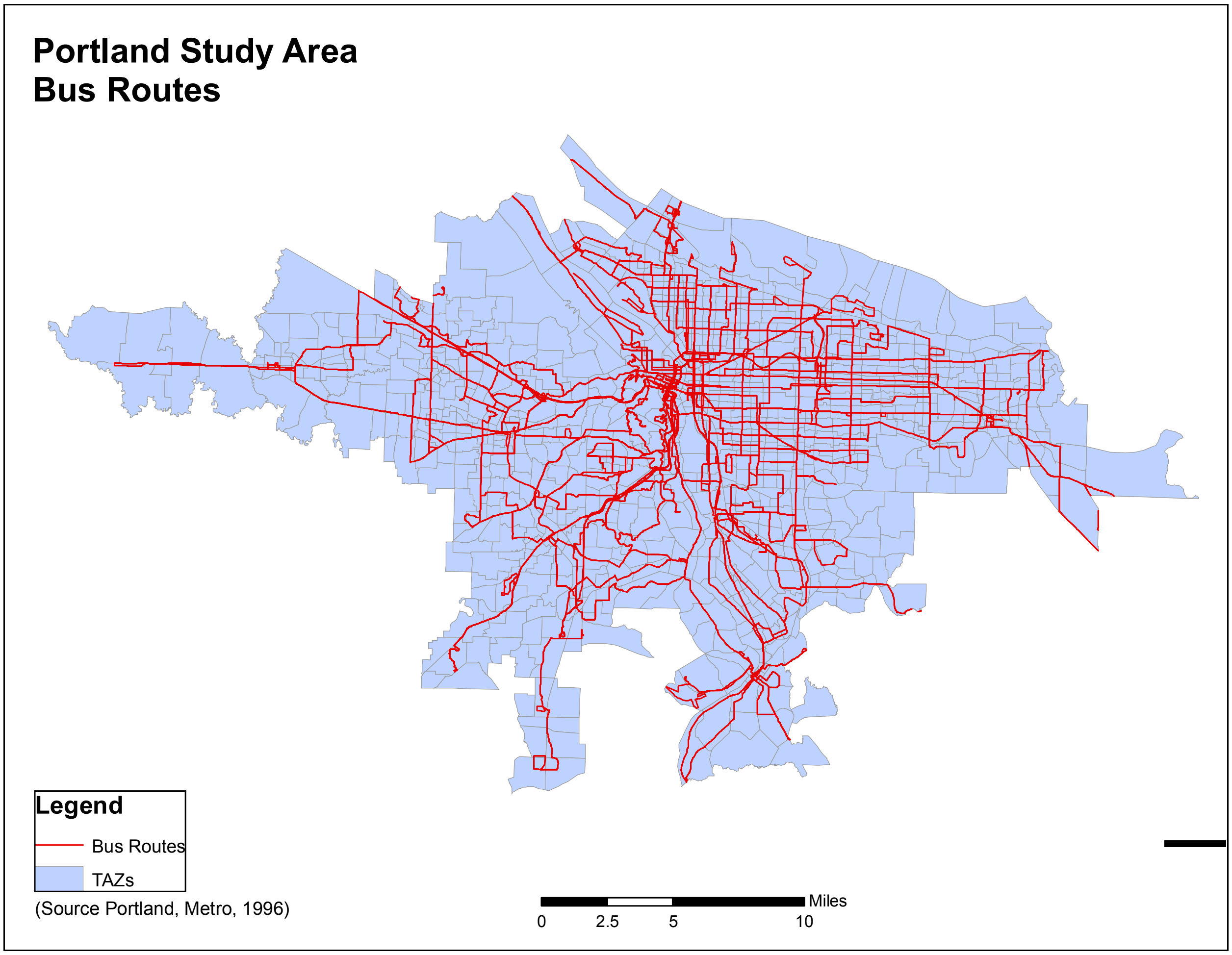


Figure 16. Commuting by Transit (Raw Data).

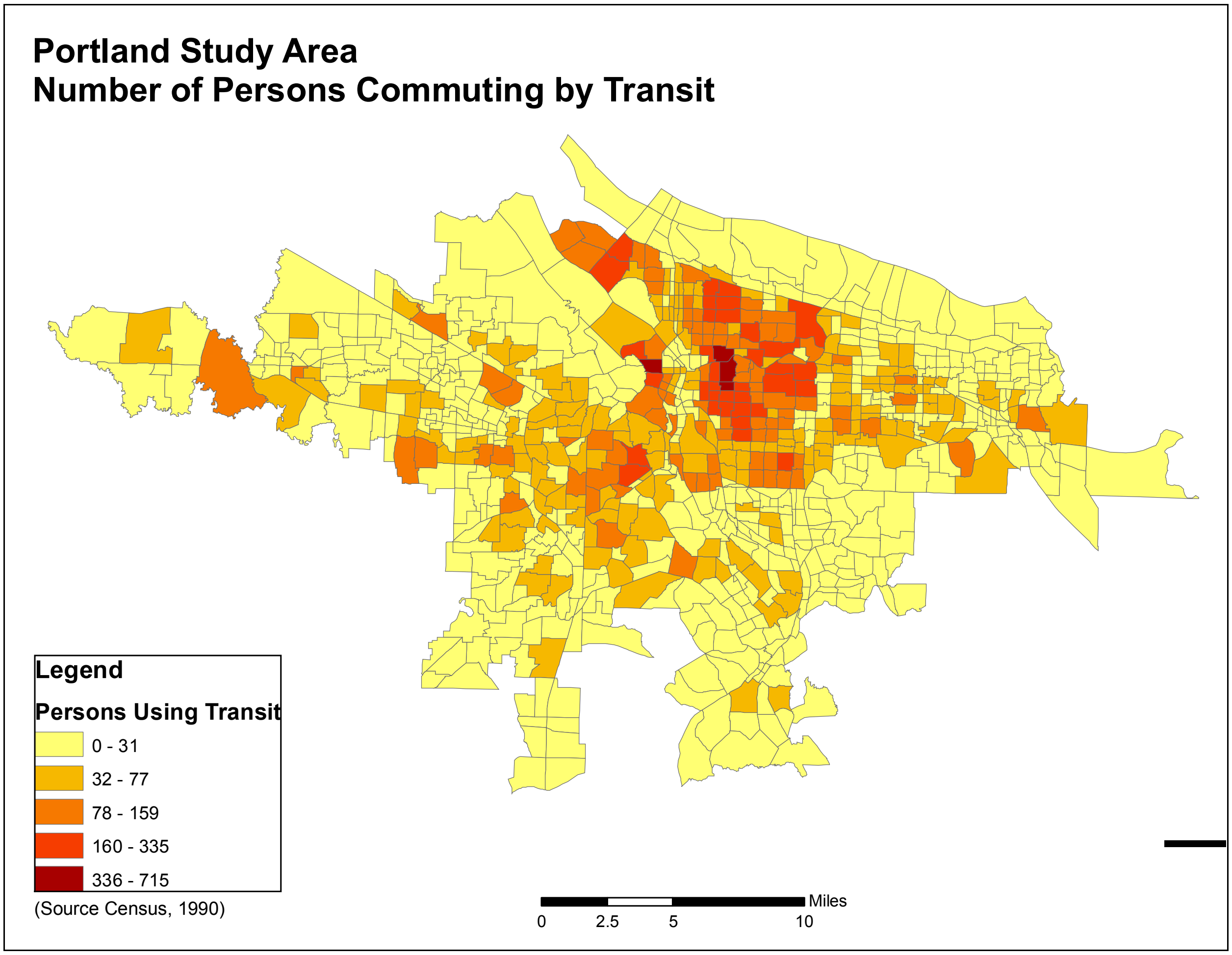


Figure 17. Light Rail Stations (Raw Data).

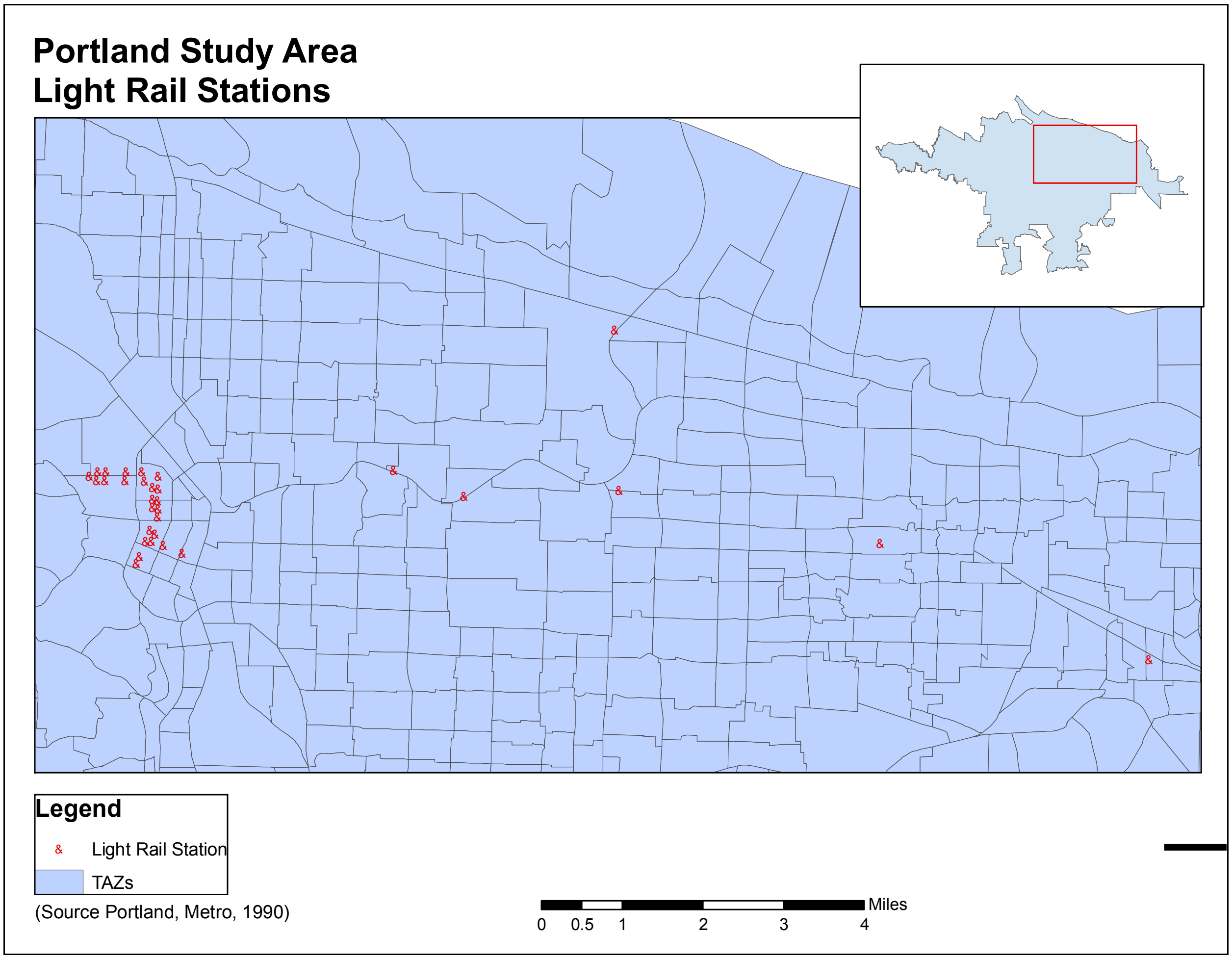


Figure 18. Streets With and Without Sidewalk Data (Raw Data).

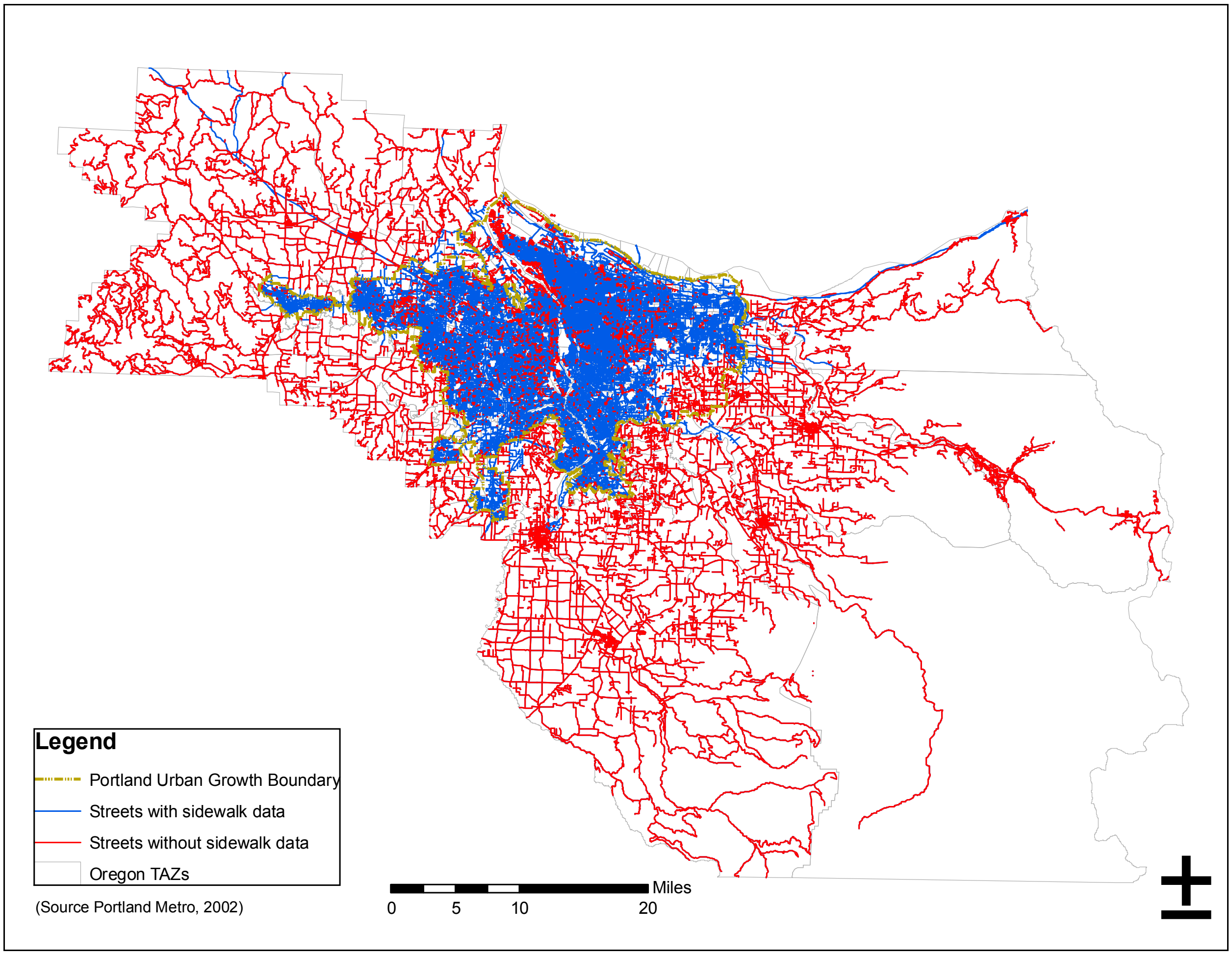


Figure 19. Commuting by Pedestrian and Bicycle Modes (Raw Data).

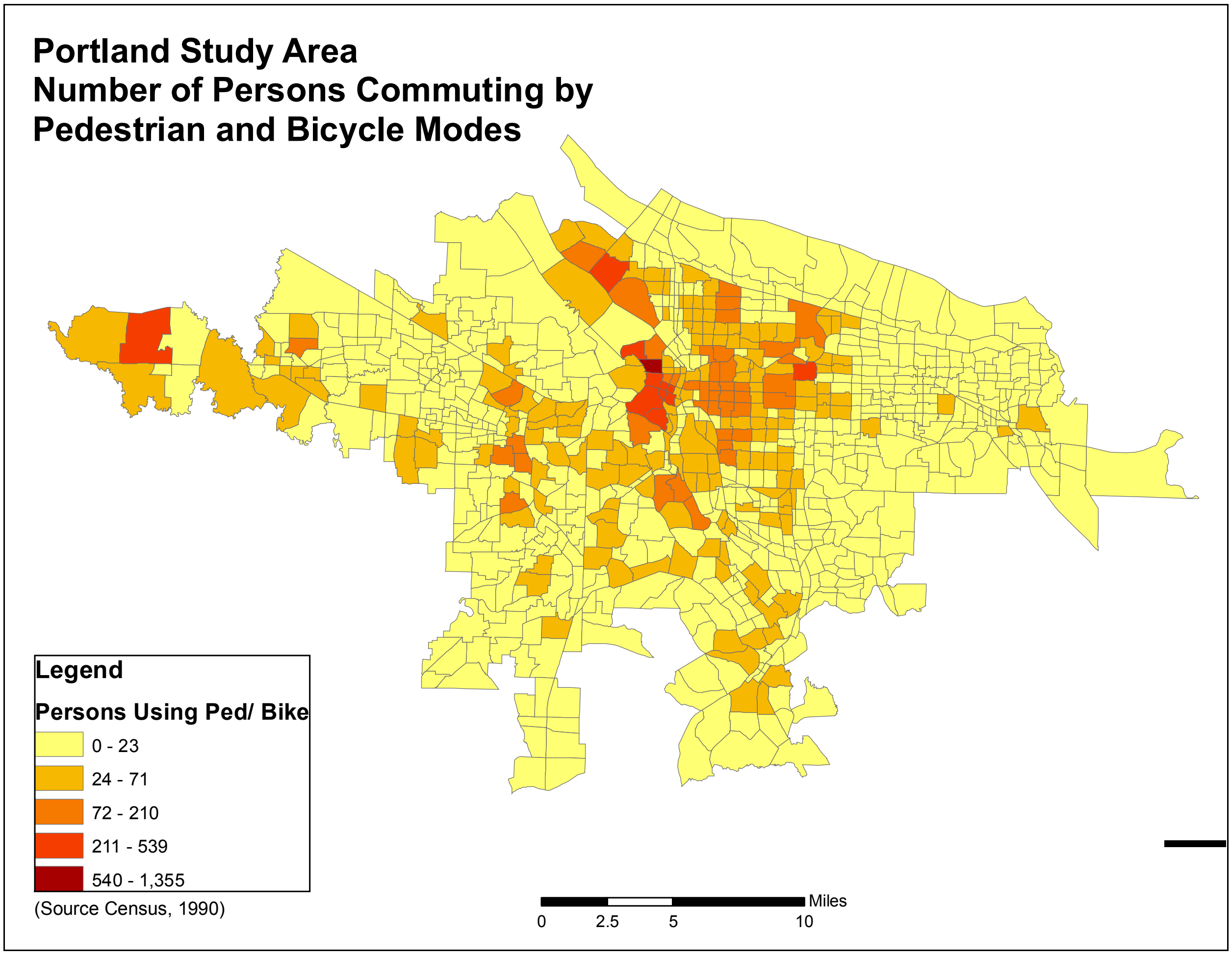


Table 7. Factor Averages for the Three Approaches.

Factors	Approach								
	Naïve			Cluster			PCA		
	urban	suburban	exurban	urban	suburban	exurban	urban	suburban	exurban
Number of TAZs	236	370	267	126	338	409	121	312	440
Development Intensity Factors									
Population density	9.38	5.33	2.40	11.32	6.55	2.91	11.35	6.86	2.99
Housing unit density	4.18	1.63	0.48	5.30	2.28	0.68	5.32	2.46	0.69
Employment density	9.86	1.80	0.64	15.45	2.30	1.07	15.98	2.62	0.94
Park density	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.00
Motorized Infrastructure Factors									
Roadway density	0.04	0.02	0.01	0.05	0.03	0.02	0.05	0.03	0.02
Bus route density	0.03	0.01	0.00	0.04	0.02	0.01	0.04	0.02	0.01
Transit commuting	12.6%	5.7%	3.4%	17.2%	6.4%	4.0%	17.3%	7.0%	3.8%
Proximity to subway station	0.58	0.00	0.00	0.93	0.05	0.00	0.96	0.06	0.00
Ped/ Bike Infrastructure Factors									
Sidewalk density	0.06	0.02	0.01	0.08	0.03	0.01	0.08	0.03	0.01
Sidewalk coverage	1.51	0.92	0.44	1.67	1.14	0.53	1.69	1.11	0.59
Ped and bicycle commuting	8.8%	3.2%	2.9%	11.4%	3.4%	3.5%	11.7%	4.0%	3.1%

This Page Intentionally Left Blank.